Journal of Applied Psychology

KENNETH E. CLARK, Editor
University of Colorado

Table of Contents

Job Attitudes in Management: I. Perceived Deficiencies in Need Fulfillment as a Function of Job Level: Lyman W. Porter 375

The Prediction of Academic Achievement with the California Psychological Inventory: Leon A. Rosenberg, Thomas B. McHenry, Anna Maria Rosenberg, and Robert C. Nichols 385

Effects of Branching in a Computer Controlled Autoinstructional Device: John E. Coulson, Donald P. Estavan, Ralph J. Melaragno, and Harry F. Silberman 389

Deceptive Packaging: Are the Deceivers Being Deceived?: James C. Naylor 393

Class Assignment in the Latent Profile Model: W. A. Gibson 399

Productivity and Errors in Two Keying Tasks: A Field Study: E. T. Klemmer and G. R. Lockhead 401

The Validity of Two-Scale Pattern Interpretation on the California Psychological Inventory: Alfred B. Heilbrun, Jr., John L. Daniel, Leonard D. Goodstein, Richard R. Stephenson, and John O. Crites 409

Re-evaluation of a Weighted Application Blank for Office Personnel: Paul F. Wernimont 417

Motion Prediction as a Function of Target Speed and Duration of Presentation: Earl L. Wiener 420

Vigilance Performance as a Function of Interpolated Rest: Bruce O. Bergum and Donald J. Lehr 425

The Identification of Traits Related to Managerial Success: John A. Hicks and Joics B. Stone 428

This is the last issue of Volume 46.
Volume Title Page and Contents appear herein.

American Psychological Association

Volume 46, Number 6 December 1962

Consulting Editors

ARTHUR BRAYFIELD, *American Psychological Association*
GEORGE E. BRIGGS, *Ohio State University*
NORMAN FREDERIKSEN, *Educational Testing Service*
LEONARD D. GOODSTEIN, *University of Iowa*
EDWIN R. HENRY, *Standard Oil Company of New Jersey*
JOHN HOLLAND, *National Merit Scholarship Corporation*
CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*
LAURENCE S. MCGAUGHRAN, *University of Houston*
QUINN MCNEMAR, *Stanford University*
HAROLD F. ROTHE, *Beloit Corporation*
THOMAS A. RYAN, *Cornell University*
CLARK L. WILSON, *Batten, Barton, Durstine, and Osborn, Incorporated*

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: vocational and educational prognosis, diagnosis, and guidance at the secondary and college level; personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research.

Manuscripts must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or they may be obtained from the Editor or from the APA Central Office.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark
Office of the Dean
College of Arts and Sciences
University of Colorado
Boulder, Colorado

All manuscripts must be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographs or pencil-drawn copies.

Manuscripts must conform to the style requirements described in the *Publication Manual of the American Psychological Association.*

Journal of Applied Psychology

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W.
Washington 6, D. C.

$10.00 per volume $2.00 per issue

HELEN ORR, Managing Editor
ELIZABETH S. REED, Advertising Manager
VIRGINIA RICHARDS, Editorial Assistant

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the subscription office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second class postage paid at Lancaster, Pennsylvania and at additional mailing places.

© 1962 by the American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 46, NO. 6 DECEMBER 1962

JOB ATTITUDES IN MANAGEMENT:

I. PERCEIVED DEFICIENCIES IN NEED FULFILLMENT AS A FUNCTION OF JOB LEVEL [1]

LYMAN W. PORTER

University of California, Berkeley

This questionnaire study investigated perceived need fulfillment deficiencies in nearly 2000 managerial positions. Respondents represented all levels of management and a wide variety of companies. 5 need categories, chosen to represent a hierarchy of prepotency of needs, were studied. Results showed: (a) Vertical level of position within management had a strong relation to the degree of perceived satisfaction of the 3 highest-order needs—Self-actualization, Autonomy, and Esteem; for these needs, satisfactions increased at each higher level of management. (b) For the 2 lower-order types of needs, Security and Social, there were no systematic changes in satisfaction in relation to management level. (c) Among the 5 need categories, Self-actualization and Autonomy were consistently regarded as the least fulfilled needs at all levels of management.

Until recently, industrial psychologists studying job attitudes have tended to focus primarily on the nonmanagement, line worker. However, in the last few years several studies have been published which reflect an increasing interest in the job attitudes of management personnel (Harrison, 1960; Herzberg, Mausner, & Snyderman, 1959; Porter, 1961a, 1961b; Rosen, 1961a, 1961b; Rosen & Weaver, 1960; Triandis, 1959a, 1959b). Most of these recent studies, especially those of Rosen and Porter, have paid particular attention to the hierarchical structure of management as a factor influencing job perceptions. This variable, level of position with the hierarchy, is the basic independent variable studied in the present investigation.

The present study is designed to extend and enlarge upon the results of a previous study of perceived need satisfactions in management jobs (Porter, 1961b). This earlier study utilized a modified Maslow-type categorization of needs in investigating perceived deficiencies in fulfillment and perceived importance of various needs. The five areas of needs studied, from lowest-order (most basic or prepotent) to highest-order (least basic or prepotent), were security, social, esteem, autonomy, and self-actualization needs. Questionnaire respondents in this previous study were drawn from the lowest levels of management in three different organizations, and the data analysis was focused on differences in perceptions between individuals in first-level supervisory jobs versus those in lower-middle management jobs. The results of this initial study showed that level of position (at least at these lower levels of management) was related to the amount of perceived deficiencies in need fulfillment, with the lowest level indicating larger deficiencies than the level above them. Among the five need categories, the highest-order areas of self-actualization and

[1] This study was carried out as part of the research program of the Institute of Industrial Relations, University of California, Berkeley. It was started while the author was a Ford Foundation Faculty Research Fellow. The Institute of Social Sciences at the University of California and the American Management Association contributed to the support of the research assistance.

The author is indebted to Mildred Henry, Larry Stewart, and Robert Andrews for assistance in tabulation of the data.

autonomy were regarded as the least satisfied needs.

The purpose of the present study is to investigate differences in perceived deficiencies in need fulfillment at all levels of management from the first level of supervision to the presidential level. The present study extends the previous study by providing data from a sample that has a considerably larger N, that is drawn from numerous types and sizes of companies located throughout the country, and, especially, that represents management from the bottom to the very top.

METHOD

Questionnaire

As in the previous study (Porter, 1961b), the data for this study were obtained by means of a questionnaire. This questionnaire contained three sections, two of which will not be reported further here. The relevant part of the questionnaire contained 13 items classifiable into a Maslow-type need hierarchy system. The printed instructions for this part of the questionnaire stated:

> On the following pages will be listed several characteristics or qualities connected with your management position. For each such characteristic, you will be asked to give three ratings:
>
> a. *How much* of the characteristic *is there now* connected with your management position?
>
> b. *How much* of the characteristic do you think *should be* connected with your management position?
>
> c. *How important* is this position characteristic *to you?*

For each of the 13 items, the respondents were instructed to answer the above three questions by circling a number on a rating scale extending from 1 to 7, where "low numbers represent low or minimum amounts, and high numbers represent high or maximum amounts." A typical item appeared as follows on the questionnaire:

The opportunity for independent thought and action in my management position:

(a) How much is there now?
(min) 1 2 3 4 5 6 7 (max)
(b) How much should there be?
(min) 1 2 3 4 5 6 7 (max)
(c) How important is this to me?
(min) 1 2 3 4 5 6 7 (max)

As will be explained in the Results section, only the answers to Parts a and b of each item were used to assess the degree of need deficiency. Answers to Part c on "importance" will be reported in a future paper.

Categories of Needs and Specific Items

Listed below are the categories of needs studied in this investigation, along with the specific items used to elicit information on each category. The items were randomly presented in the questionnaire, but are here listed systematically according to their respective need categories. The rationale behind the categorization system has been presented in the previous article (Porter, 1961b) and will not be repeated here. It can be pointed out, though, that the system is designed to be in general agreement with Maslow's hierarchical classification scheme based on the supposed "prepotency" (or priority) of various types of needs (Maslow, 1954). The categories, arranged from lowest-order (most prepotent) to highest-order (least prepotent), and their specific items follow:

I. Security needs
1. The *feeling of security* in my management position

II. Social needs
1. The *opportunity*, in my management position, *to give help to other people*
2. The *opportunity to develop close friendships* in my management position

III. Esteem needs
1. The *feeling of self-esteem* a person gets from being in my management position
2. The *prestige* of my management position *inside* the company (that is, the regard received from others *in* the company)
3. The *prestige* of my management position *outside* the company (that is, the regard received from others *not* in the company)

IV. Autonomy needs
1. The *authority* connected with my management position
2. The *opportunity for independent thought and action* in my management position
3. The *opportunity*, in my management position, for *participation in the setting of goals*
4. The *opportunity*, in my management position, *for participation in the determination of methods and procedures*

V. Self-actualization needs
1. The *opportunity for personal growth and development* in my management position
2. The *feeling of self-fulfillment* a person gets from being in my management position (that is, the feeling of being able to use one's own unique capabilities, realizing one's potentialities)
3. The *feeling of worthwhile accomplishment* in my management position

Procedure and Sample

The questionnaire was distributed to approximately 6,000 managers and executives throughout the country. It was sent to a random sample of 3,000 members (approximately 10%) of the American Management Association and to another random sample of some 3,000 managers whose names were

TABLE 1

DISTRIBUTION OF *N* OF TOTAL SAMPLE BY FIVE MANAGEMENT LEVELS AND FOUR AGE GROUPS, AND CHARACTERISTICS OF SAMPLE BY MANAGEMENT LEVEL

Management level	Age group				Total *N* for level	Median age	College degree (%)
	20–34	35–44	45–54	55+			
President	7	31	51	25	114	48.2	76.1
Vice President	52	240	212	107	611	45.5	72.6
Upper-Middle	95	288	206	70	659	43.1	75.0
Lower-Middle	100	208	98	25	431	40.6	75.0
Lower	32	46	14	9	101	39.2	76.2

on mailing lists available to the Association.[2] The distribution of the questionnaire was accomplished by mail, with enclosures including a covering letter from the Association and a prestamped return envelope, as well as the 12-page questionnaire. Responses were received from 1,958 managers (for a response rate of approximately 33%), with the number of usable questionnaires being 1,916.

From personal data questions asked on the last page of the questionnaire, it was possible to classify respondents on a number of independent variables. The two relevant variables for this study were level of position within management and (as a control variable) age of the respondent. For level of position, a five-category system was set up. These categories were (from top to bottom):

President: presidents and chairmen of boards
Vice President: vice presidents (or their equivalents in large companies)
Upper-Middle: approximately the level of division managers, plant managers, and major department managers
Lower-Middle: approximately the level of department and subdepartment managers
Lower: first-level or second-level supervisors

Assignment of respondents to the above categories of level of position was made on the following basis: Anyone having the title of president or chairman of a board of directors was automatically put in the President category, and this category was limited to these two chief executive positions. (Presidents accounted for 90% of the individuals in this category.) Individuals were assigned to the Vice President category on the basis of having a vice president title (including such titles as assistant vice president) or its equivalent in very large companies. (Examples of equivalent types of positions would be certain staff positions, such as controller. Thus, although the title was not that of a vice president, the respondent would be put in that category if he indicated that he reported to the chief executive of the organization.) Respondents who did not qualify for the President or Vice President categories were assigned to one of the other three levels on the basis of their answers to two questions: (*a*) How many levels of supervision are there in your organization? (*b*) How many levels of supervision are there *above* your position? Those respondents who indicated by their answers to these two questions that they were in the lowest level of supervision were automatically placed in the Lower (bottom) category. The remaining respondents were placed in the Upper-Middle, Lower-Middle, or Lower category, depending upon the ratio of levels of supervision above them to the total levels of supervision in their organizations. A system of "fractions" was worked out for this assignment. The key feature of this fraction system was that respondents were classified together in the same category based on their relative distance up the management hierarchy, regardless of size of company.

For this study, level of position was the major independent variable. However, since this variable is to some extent correlated with the age of the respondent, it was felt necessary to classify respondents by age as well as by level of position. In this way, by looking at differences among management levels within different age groups, the age variable could be more or less held constant and the level variable independently assessed. The data for classification by age were obtained from one of the personal data questions that asked the respondent to indicate his age in one of eight age categories. In the data analysis, the five management levels were cross-tabulated against four age groups: 20–34, 35–44, 45–54, and 55 +. A further important advantage of tabulating responses by age as well as by management level was that by having four age groups of five management levels, it was possible to replicate management level effects in four independent samples.

Table 1 presents the characteristics of the sample by the five management levels. This table shows that, as would be expected, the higher levels have somewhat older personnel. With regard to amount of formal education, however, the five levels are approximately equivalent in terms of the percentage of respondents holding college degrees. This means that the lowest level in this sample is probably

[2] The assistance of the American Management Association, and particularly Robert F. Steadman, in obtaining the sample of respondents is gratefully acknowledged.

composed mostly of office-type supervisory and staff personnel rather than foremen of blue collar workers because of its relatively high percentage of college educated respondents. Therefore, any conclusions involving the lowest level of this sample would not necessarily apply to first-line foremen who do not have as high a level of formal schooling. However, the fact that all five levels in the sample for this study have almost identical amounts of formal education makes it possible to assess trends for effects of level independently of those for amount of education.

Table 1 also presents the number of respondents in each management level within each age group. This table can be referred to in determining the *N*s for the various cells in succeeding tables.

One other relevant distribution of the sample (not shown in Table 1) is that for type of company. About 66% of the sample came from manufacturing companies, 7% from transportation and public utilities, 7% from finance and insurance, 5% from wholesale and retail trade, and the remaining 15% was from among other types of companies. It should be stressed that due to the method of distribution of the questionnaire, a nationwide sample was obtained; furthermore, except by chance, any particular company would not be represented more than a few times in the total sample.

RESULTS

The degree of perceived deficiency in fulfillment for each respondent on each questionnaire item was obtained by subtracting the answer to Part a of an item ("How much of the characteristic is *now connected* with your position?") from Part b of the item ("How much of the characteristic do you think *should be connected* with your position?"). An a priori assumption was made that the larger the difference—a subtracted from b—the larger the degree of dissatisfaction or the smaller the degree of satisfaction. This method of measuring perceived need satisfaction thus is an indirect measure derived from two direct answers by the respondent for each item. This method has two presumed advantages: (*a*) The subject is not asked directly concerning his satisfaction. Therefore, any tendency for a simple "response set" to determine his expression of satisfaction is probably reduced somewhat. It is more difficult, although by no means impossible, for the respondent to manipulate his satisfaction measure to conform to what he thinks he "ought" to put down versus what he actually feels to be the real situation. (*b*) Secondly, this method of measuring need fulfillment is a more conservative measure than would be a single question concerning simple obtained satisfaction. It takes into account the fact that higher level positions should be expected to provide more rewards because it utilizes the difference between obtained and expected satisfaction. In effect, this method asks the respondent, "how satisfied are you in terms of what you expected from this particular management position?" Thus, it is designed to be a realistic and meaningful measure in comparing different management groups.

Table 2 presents the mean need fulfillment deficiencies for each of the 13 items in the questionnaire and for each subgroup of respondents. The values for each entry in Table 2 were obtained by subtracting each respondent's answer to Part a of each item from his answer to Part b of each item, and then calculating the mean of these values for each subgroup of respondents. No entries are given where the *N* for a cell is less than 20 because of the lack of stability in values based on such small numbers of respondents.

Examination of Table 2 shows several general trends that will be brought out more clearly in succeeding tables. First, Table 2 shows that need fulfillment deficiencies for most items definitely tended to increase at each successive lower level of the management hierarchy. This occurred within all four age groups. The second major trend observable in this table is that the largest deficiencies were found in the items in the two highest-order need categories, Autonomy and Self-Actualization.

Table 3 brings into sharper focus the first trend apparent in Table 2, the trend for larger need fulfillment deficiencies to be associated with the lower levels of management. Table 3 presents data on the number of changes in size of mean deficiencies from the higher to the lower levels of management. The entries for this table were obtained in the following manner: Changes of mean deficiencies greater than .05 scale units from one management level to the next in Table 2 were counted as increases (+s) or decreases (−s) in deficiencies, depending on the direction of change. Increases or decreases of .05 or less scale units were counted as "no

TABLE 2

MEAN NEED FULFILLMENT DEFICIENCIES FOR EACH NEED CATEGORY AND ITEM: FIVE MANAGEMENT LEVELS by FOUR AGE GROUPS

Need category	Item	Age group	Management level				
			President	Vice President	Upper-Middle	Lower-Middle	Lower
Security	I-1	20–34	—	0.15	0.05	0.32	0.19
		35–44	0.34	0.51	0.44	0.27	1.22
		45–54	0.33	0.43	0.46	0.68	—
		55+	−0.24	0.49	0.57	0.36	—
Social	II-1	20–34	—	0.48	0.39	0.52	0.50
		35–44	0.56	0.31	0.49	0.37	0.87
		45–54	0.24	0.49	0.42	0.38	—
		55+	0.32	0.43	0.44	0.40	—
	II-2	20–34	—	0.21	0.19	0.12	0.06
		35–44	0.37	0.08	0.24	0.26	0.80
		45–54	0.45	0.30	0.20	0.34	—
		55+	0.24	0.14	0.20	0.20	—
Esteem	III-1	20–34	—	0.42	0.79	1.11	1.44
		35–44	0.50	0.64	0.93	0.89	1.41
		45–54	0.41	0.57	0.85	0.81	—
		55+	0.16	0.69	0.74	0.76	—
	III-2	20–34	—	0.38	0.62	0.94	1.31
		35–44	0.03	0.41	0.71	0.79	1.52
		45–54	0.39	0.40	0.73	0.99	—
		55+	0.20	0.46	0.59	0.80	—
	III-3	20–34	—	0.33	0.46	0.25	0.81
		35–44	0.31	0.43	0.48	0.45	0.74
		45–54	0.27	0.29	0.39	0.27	—
		55+	0.20	0.23	0.16	0.16	—
Autonomy	IV-1	20–34	—	0.56	0.81	1.22	1.19
		35–44	0.09	0.65	0.96	0.92	1.67
		45–54	0.35	0.67	0.93	0.86	—
		55+	0.36	0.69	0.69	1.00	—
	IV-2	20–34	—	0.42	0.77	0.85	1.50
		35–44	0.34	0.52	0.76	0.84	1.41
		45–54	0.29	0.44	0.65	0.84	—
		55+	0.28	0.57	0.70	1.00	—
	IV-3	20–34	—	0.42	1.16	1.40	1.56
		35–44	0.53	0.62	1.17	1.26	1.96
		45–54	0.20	0.83	1.17	1.16	—
		55+	0.24	0.58	1.03	1.32	—

Table 2—Continued

Need category	Item	Age group	Management level: President	Vice President	Upper-Middle	Lower-Middle	Lower
	IV-4	20–34	—	0.21	0.69	0.67	0.94
		35–44	−0.16	0.33	0.71	0.59	1.35
		45–54	−0.14	0.50	0.69	0.88	—
		55+	0.04	0.46	0.90	1.28	—
Self-Actualization	V-1	20–34	—	0.71	0.98	1.01	1.19
		35–44	0.44	0.91	0.97	1.13	2.02
		45–54	0.57	0.97	1.17	1.15	—
		55+	0.52	0.93	1.36	1.20	—
	V-2	20–34	—	0.77	1.11	1.17	1.53
		35–44	0.44	0.90	1.18	1.28	1.87
		45–54	0.65	0.80	1.10	1.00	—
		55+	0.44	0.87	0.84	0.84	—
	V-3	20–34	—	0.85	1.04	1.37	1.50
		35–44	1.03	1.00	1.26	1.20	1.83
		45–54	0.90	0.99	1.19	1.15	—
		55+	0.76	0.79	1.04	1.08	—

Note.—Measures of variability are not included in this table. However, the standard deviations for each mean in the table were computed and they showed the following distribution by quartiles: $Q_1 = 1.00$, $Q_2 = 1.15$, $Q_3 = 1.33$.

changes" (0s). (The decision to use .05 scale unit as a dividing line between "changes" and no changes was arbitrary. The use of the .05 separation allowed for trends to show in the data while at the same time reducing the opportunity for some chance fluctuations in sizes of means to affect the trends.) In Table 3 the number of each type of changes—increases, decreases, or no changes—in size of mean deficiencies from higher to lower management levels were recorded in the columns for each age group. Thus, for example, for Esteem Item III-1 ("The feeling of self-esteem a person gets from being in my management position"), Table 3 shows that in the 20–34 age group the mean deficiency increased three times in going from the top to the bottom management levels; in the 35–44 age group, there were three increases and one no change; and in the 45–54 age group there were two increases and one no change; and in the 55+ age group there were one increase and two no changes. Combining the tabulations for all four age groups for this item shows that there were nine instances of increase, four of no change, and no instances of decrease. Such a result is, by a simple sign test, significant at the .01 level of confidence.

The major results of the study are presented in the last set of three columns in Table 3. The three columns for the total sample show the total number of increases, decreases, and no changes for each of the 13 items. These columns also summarize the results for each of the five need categories by totaling the item results within each category. The p values for each item and category are listed in the footnotes of Table 3.

Table 3 shows conclusively that need fulfillment deficiencies progressively increased from the top to the bottom of the management hierarchy for three of the five need categories—Esteem, Autonomy, and Self-Actualization. Eight of the 10 specific items in these three categories showed significant trends of increase in deficiencies at or exceeding the .05 level of confidence (by a sign test), one item showed a trend significant at the .10 level of confidence, and only one of the 10 items failed to produce a trend approaching statistical significance. On the other hand, for

the other two categories—Security and Social needs—there were no significant trends within items or categories. Thus in the two lowest-order need areas, security and social satisfactions, lower-level managers perceived themselves to be about as satisfied as higher-level managers. However, for the three highest-order need areas—Esteem, Autonomy, and Self-Actualization—lower-level managers perceived their positions as providing significantly less satisfaction compared with higher-level managers. It is especially interesting to note that these overall findings were supported with regularity in each of the four age groups, showing that these trends are not merely a function of higher-level managers being older, on the average, than lower-level managers. Age differences among managerial levels clearly do not explain the differences in perceived need fulfillment deficiencies that occurred among the levels.

In Table 4, the rank orders of the five need categories are presented. These rank orders are in terms of the size of the mean need fulfillment deficiencies for each category of items for each subgroup of respondents. The category values on which these rankings are based were calculated on the basis of the mean of the item deficiencies within each category. For example, at the Vice President level and within the 35–44 age group, the Self-Actualization category had the largest mean deficiency, followed by the Autonomy, Security, Esteem, and Social categories.

The outstanding feature of Table 4 is that the Self-Actualization and Autonomy need areas consistently ranked first and second, respectively, in terms of size of mean deficien-

TABLE 3

NUMBER OF CHANGES IN SIZE OF MEAN DEFICIENCIES FROM HIGHER TO LOWER LEVELS OF MANAGEMENT WITHIN FOUR AGE GROUPS

Need category	Item	Age group 20–34			35–44			45–54			55+			Total sample		
		+	0	−	+	0	−	+	0	−	+	0	−	+	0	−
Security	I-1	1	0	2	2	0	2	2	1	0	2	0	1	7	1	5
Social	II-1	1	1	1	2	0	2	1	1	1	1	2	0	5	4	4
	II-2	0	1	2	2	1	1	1	0	2	1	1	1	4	3	6
Category total		1	2	3	4	1	3	2	1	3	2	3	1	9	7	10
Esteem	III-1	3	0	0	3	1	0	2	1	0	1	2	0	9	4	0**
	III-2	3	0	0	4	0	0	2	1	0	3	0	0	12	1	0**
	III-3	2	0	1	2	2	0	1	1	1	0	2	1	5	5	3
Category total		8	0	1	9	3	0	5	3	1	4	4	1	26	10	3**
Autonomy	IV-1	2	1	0	3	1	0	2	0	1	2	1	0	9	3	1*
	IV-2	3	0	0	4	0	0	3	0	0	3	0	0	13	0	0**
	IV-3	3	0	0	4	0	0	2	1	0	3	0	0	12	1	0**
	IV-4	2	1	0	3	0	1	3	0	0	3	0	0	11	1	1**
Category total		10	2	0	14	1	1	10	1	1	11	1	0	45	5	2**
Self-Actualization	V-1	2	1	0	4	0	0	2	1	0	2	0	1	10	2	1*
	V-2	3	0	0	4	0	0	2	0	1	1	2	0	10	2	1*
	V-3	3	0	0	2	1	1	2	1	0	1	2	0	8	4	1*
Category total		8	1	0	10	1	1	6	2	1	4	4	1	28	8	3**
Total for all items		28	5	6	39	6	7	25	8	6	23	12	4	115	31	23**

* Approaches significance ($p = .10$).
** $p = .05$.
*** $p = .01$.

TABLE 4

RANKS OF MEAN DEFICIENCIES FOR FIVE NEED CATEGORIES WITHIN SUBGROUPS OF RESPONDENTS: FIVE MANAGEMENT LEVELS by FOUR AGE GROUPS

Management level	Age group	Need category				
		Security	Social	Esteem	Autonomy	Self-Actualization
President	20–34	—	—	—	—	—
	35–44	3	2	4	5	1
	45–54	4	3	2	5	1
	55+	5	2	4	3	1
Vice President	20–34	5	4	3	2	1
	35–44	3	5	4	2	1
	45–54	3	5	4	2	1
	55+	3	5	4	2	1
Upper-Middle	20–34	5	4	3	2	1
	35–44	4	5	3	2	1
	45–54	4	5	3	2	1
	55+	3	5	4	2	1
Lower-Middle	20–34	4.5	4.5	3	2	1
	35–44	5	4	3	2	1
	45–54	4	5	3	2	1
	55+	4	5	3	1	2
Lower	20–34	5	4	3	2	1
	35–44	3.5	5	3.5	2	1
	45–54	—	—	—	—	—
	55+	—	—	—	—	—

cies in nearly all subgroups of respondents. Among the other three need categories, Esteem more often ranked third in size of deficiencies, while the Social and Security need areas were about equal in being the areas with the smallest deficiencies. Secondly, Table 4 shows that the different managerial levels produced similar rankings of the five need areas in terms of deficiencies, although there were a few exceptions. One exception was the President level, where deficiencies in Social needs had relatively higher ranks and those for Autonomy lower ranks than was the case with the other management levels. Another exception was the tendency for deficiencies in Esteem needs to receive somewhat higher ranks at lower compared with upper levels of management. The third finding apparent in Table 4 is that deficiencies in Security needs tended to rank higher for older individuals than for younger individuals (except at the chief executive level), whereas deficiencies in Social needs tended to receive somewhat higher ranks in younger compared with older manager groups at most levels.

DISCUSSION

Before considering the implications of the results of this study, the findings can first be compared with those from two previous relevant studies that were carried out on much smaller and less representative samples. In general, the results of the present study tend to confirm the earlier study carried out with the same set of need satisfaction questions (Porter, 1961b). The earlier study, based on small samples from three companies, found consistent decreases in need fulfillment deficiencies between bottom management and the level immediately above it (equivalent to

approximately the Lower-Middle level in this study). The current study shows generally the same finding at the two lowest levels, but also completes the picture by showing that this trend continues up to the very top management levels. The two studies are also similar in showing that the two highest-order need areas, Autonomy and Self-Actualization, produce the largest deficiencies. In certain more specific aspects, the two studies show some other similarities and also some differences in findings. An example of one of the differences is that the previous investigation indicated a trend of decreasing deficiency with increasing management level in the Security need area, while the present study failed to find any meaningful trend in the Security area. Security needs seemed about as well satisfied at bottom as at top management. Another difference is that the former study showed no appreciable changes in Self-Actualization deficiencies between the two management levels studied, whereas in the present, larger investigation, there was a trend for deficiencies in this area to decrease with increasing levels of the hierarchy (although the deficiencies were still large in an absolute sense even at the top levels). Specific similarities between the two studies involved the Social need area which produced no trend in either study, and the Esteem and Autonomy areas which produced definite trends in both studies. In sum, where comparisons between the results for the previous study and present study can be made—at the Lower and Lower-Middle levels of management—the studies tend to be in general agreement. In those specific instances where the findings differ, the later study is probably more accurate because of its vastly increased size of sample, its considerable spread in types and sizes of companies represented, and because differences in amount of formal education between the two lower levels are controlled.

The results of this study can also be compared with those of a previous investigation by Rosen (1961a). The Rosen study reported on satisfactions with conditions of work given by managers from three levels (plus a group of staff managers) in a single industrial plant. Greater satisfaction with work conditions was found at higher management levels (the highest probably corresponding to the Upper-Middle level of this study), but the three levels produced similar rankings of the conditions in terms of degree of satisfaction. These results from a single specific company, although based on an entirely different set of satisfaction questions, are in agreement with the results of the current study.

The major implications of the findings of the present study of nearly 2,000 managers would appear to be two:

1. The vertical location of management positions is an important factor in determining the extent to which managers feel that they can satisfy particular psychological needs, especially the three higher-order needs of a Maslow-type system of need hierarchies. It appears that there does exist a *differential* opportunity within management to satisfy Esteem, Autonomy, and Self-Actualization needs. Thus, individuals in the top policy-making positions of organizations may find it necessary to be as concerned with the satisfactions of their lower-level managers as they are with the satisfactions of their blue collar workers who usually have a union or other collective bargaining agency to represent their interests. Although this study does not necessarily show lower- and middle-level managers to be highly dissatisfied, it does definitely show them to be more dissatisfied than managers at top-level positions. In this connection, it should also be recalled that the measure of satisfaction (or dissatisfaction) used in this study took into account the fact that higher levels of management should be expected to provide greater rewards. The increasing dissatisfaction at lower levels of management represents the increasing difference between what is expected and what is obtained. To put it another way, lower-level managers' expectations in most of the specific need satisfaction areas are much more divergent from their perception of reality than is the case for top managers. For this situation to be changed, it would seem that either lower-level managers would have to change their expectations, or upper-level managers would have to change the chances for satisfaction in lower management, especially in the highest-order need areas.

2. The Self-Actualization, Autonomy, and Esteem areas seem to be the most critical areas of need fulfillment deficiencies at all levels of management, with perhaps the possible exception of the very top, chief executive level. These three areas have always been mentioned as ones that are relatively unsatisfied at the nonmanagement worker level, but the corollary assumption has also been that these are probably well satisfied throughout management. The results of this study cast some doubt on this assumption. Certainly they show that these three need areas are not nearly so well satisfied within management as the social and security need areas. Whether, in an absolute as well as a relative sense, they show a high level of dissatisfaction in management is a somewhat harder question to answer. Should, for example, mean differences between expectation and perceived reality of around 1.00 scale units in the Upper-Middle manager level for the Self-Actualization items be considered large, medium, or small amounts of dissatisfaction? It is difficult to say. However, these particular means correspond to approximately 50–60% of individuals indicating some degree of deficiency on each item. Thus, taking all of the evidence together, it would seem reasonable to conclude that fairly large numbers of managers even at high levels of organizations are not satisfied with their opportunities to obtain the amount of Self-Actualization they think should be available from their jobs. The same conclusion would hold true for the Esteem and Autonomy need areas, although to a slightly reduced extent.

REFERENCES

HARRISON, R. L. Sources of variation in managers' job attitudes. *Personnel Psychol.*, 1960, **13**, 425–434.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, BARBARA B. *The motivation to work*. New York: Wiley, 1959.

MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.

PORTER, L. W. Perceived trait requirements in bottom and middle management jobs. *J. appl. Psychol.*, 1961, **45**, 232–236. (a)

PORTER, L. W. A study of perceived need satisfactions in bottom and middle management jobs. *J. appl. Psychol.*, 1961, **45**, 1–10. (b)

ROSEN, H. Desirable attributes of work: Four levels of management describe their job environments. *J. appl. Psychol.*, 1961, **45**, 156–160. (a)

ROSEN, H. Managerial role interaction: A study of three managerial levels. *J. appl. Psychol.*, 1961, **45**, 30–34. (b)

ROSEN, H., & WEAVER, C. G. Motivation in management: A study of four managerial levels. *J. appl. Psychol.*, 1960, **44**, 386–392.

TRIANDIS, H. C. Categories of thought of managers, clerks, and workers about jobs and people in industry. *J. appl. Psychol.*, 1959, **43**, 338–344. (a)

TRIANDIS, H. C. Differential perception of certain jobs and people by managers, clerks, and workers in industry. *J. appl. Psychol.*, 1959, **43**, 221–225. (b)

(Received March 5, 1962)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 385–388

THE PREDICTION OF ACADEMIC ACHIEVEMENT WITH THE CALIFORNIA PSYCHOLOGICAL INVENTORY

LEON A. ROSENBERG,[1] THOMAS B. McHENRY [2]
Medical Field Service School, Fort Sam Houston
ANNA MARIA ROSENBERG [3]
Community Guidance Center of Bexar County, Texas
AND ROBERT C. NICHOLS [4]
Purdue University

An attempt to use the California Psychological Inventory to predict the academic grades of students in 3 military courses: Neuropsychiatric, Clinical Psychology, and Social Work Procedures. The data indicated that a combination of the Achievement via Independence scale of the California Psychological Inventory and the General Technical score from the Army Classification Battery, makes an effective instrument for screening students for admittance to these courses. The predicted criterion scores, obtained from 2 multiple prediction equations, were correlated with the academic grades achieved by students in 2 samples ($N = 98$; $N = 64$). The obtained r's were .610 and .600 ($p < .01$). Cross-validation correlations were obtained with 2 new samples ($N = 35$; $N = 33$). The obtained r's were .631 and .558 ($p < .01$).

This is a report of an attempt to utilize the California Psychological Inventory to predict the academic success or failure of enlisted personnel in three areas of advanced military schooling. The three courses are as follows:

1. Neuropsychiatric Procedures (NP)—a 4-week course to train enlisted personnel so that they have sufficient knowledge to assist in the care of patients with neuropsychiatric disorders. Patient management, security, and rehabilitation are discussed in detail, together with many other typical ward problems.

2. Clinical Psychology Procedures (CP)—an 8-week course to prepare students to assist the clinical psychologist by administering and scoring the Wechsler Adult Intelligence Scale, Minnesota Multiphasic Personality Inventory, and various other standarized tests.

3. Social Work Procedures (SW)—an 8-week course to train students to assist the social work officer by abstracting medical records, obtaining social histories from patients, and in carrying out treatment plans formulated, and supervised, by the professional staff.

In all three courses, stress is placed on understanding the basic concepts of normal personality development. Students are given additional instruction in the symptoms and classification of psychiatric illness and attend weekly conferences where various neuropsychiatric cases are presented and discussed.

In an effort to keep student attrition to a practicable minimum yet maintain high standards of proficiency, an attempt was made to utilize personality measures for screening purposes. An earlier study involved the use of the Minnesota Multiphasic Personality Inventory (MMPI). The results of this earlier study indicated that the MMPI could not be used for this purpose (Hewitt & Rosenberg, 1962). The present study investigated the predictive value of the California Psychological Inventory (CPI) in the same setting.

METHOD

Subjects

Two samples of subjects (*Ss*) were used. The first sample of 162 students was made up of all the students in two CP classes, three SW classes, and

[1] Now with the Esther Loring Richards Children's Center, Owings Mills, Maryland.

[2] Now at Purdue University, West Lafayette, Indiana.

[3] Now with the Baltimore County School System, Garrison, Maryland.

[4] Now with the National Scholarship Corporation, Evanston, Illinois.

TABLE 1

DESCRIPTIVE DATA FOR EACH GROUP OF SUBJECTS

	First sample			Validation sample		
	CP	SW	NP	CP	SW	NP
Age	22.96	24.12	20.90	23.23	25.21	20.72
	(6.99)	(4.94)	(2.74)	(5.83)	(5.25)	(2.50)
GT	122.77	123.47	121.28	123.57	124.71	121.93
	(11.65)	(13.16)	(12.17)	(10.54)	(12.00)	(13.27)
Education	13.28	14.12	12.90	13.38	12.78	13.00
	(1.58)	(1.80)	(1.49)	(2.26)	(1.84)	(2.47)
N	58[a]	40[b]	64[c]	21[b]	14[a]	33[b]

Note.—Mean age, GT, and educational level with standard deviations in parentheses.
[a] One female included.
[b] All males.
[c] 15 females included.

two NP classes. The cross-validation group consisted of 68 *Ss* from three new classes (1 CP, 1 SW, and 1 NP). In all, the study utilized all of the classes which were in session during the period of August 1960 to March 1961.

Table 1 presents a basic description of the *Ss* involved in the study. The requirements for entrance into the three courses are approximately the same; hence, the several groups are generally quite similar. Table 1 presents the means (and standard devia-

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR EACH CPI SCALE AND THE GT MEASURE AND THEIR CORRELATIONS WITH ACADEMIC GRADES

	Classes					
	CP + SW (*N* = 98)			NP (*N* = 64)		
Scale	*M*	*SD*	*r*	*M*	*SD*	*r*
General Technical	122.46	12.26	.577**	121.28	12.17	.549**
Dominance	29.13	6.78	.200*	28.32	5.82	.307*
Capacity for Status	21.03	3.99	.376**	20.73	3.69	.439**
Sociability	26.66	5.31	.110	27.17	4.13	.117
Social Presence	36.25	6.70	.321**	36.51	6.87	.213
Self-Acceptance	22.67	3.84	.347**	21.45	4.24	.168
Sense of Well-Being	37.41	4.89	.236*	39.18	3.83	.192
Responsibility	31.11	4.67	.248*	32.45	4.15	.206
Socialization	36.96	5.35	.069	37.34	5.27	−.057
Self-Control	29.64	7.75	−.016	32.09	7.38	−.004
Tolerance	23.04	5.66	.424**	24.00	6.51	.234*
Good Impression	19.97	6.56	−.018	21.92	6.29	−.056
Communality	25.89	1.84	.216*	26.15	1.57	−.102
Achievement via Conformance	28.29	5.37	.060	28.92	4.68	.199
Achievement via Independence	19.54	4.74	.459**	20.73	4.31	.472**
Intellectual Efficiency	39.18	6.29	.368**	41.20	3.88	.310*
Psychological-Mindedness	11.45	2.87	.210*	11.59	2.59	.204
Flexibility	9.87	4.25	.381**	10.62	4.23	.343**
Femininity	16.95	4.34	−.085	18.42	3.72	−.032

* $p < .05$.
** $p < .01$.

tions in parentheses) of the ages, General Technical scores (GT),[5] and years of education, for the *Ss* in each course of instruction.

Procedure

At the start of each course the CPI was administered to each class as a group. The tests were not scored or examined in any way until the class had graduated and final grades were available. The GT scores were obtained from the students' military records. The criteria of academic success (the students' final grades) were the averages of their grades on several objective tests. These tests covered the lecture material presented in each course and were not in any way influenced by the present researchers.

RESULTS

Table 2 presents the means and standard deviations for each CPI scale and the GT measure. In Table 2 (and for all future computations) the CP and SW classes have been combined while the NP class has been treated separately. This was done because the CP and SW courses are very similar in terms of course content, difficulty level, and course length. The NP course, on the other hand, is shorter in length and simpler in content. The CP and SW classes were also quite similar in terms of the final grades obtained in both courses. The CP class achieved an average grade of 82.14 ($SD = 8.56$) and the SW class an average grade of 83.18 ($SD = 7.48$). The NP class' grades were somewhat lower, 80.32 ($SD = 9.27$).

Table 2 also presents the obtained correlations (Pearson r) of CPI scales and GT with academic grades for the first sample.

In order to refine the predictive accuracy of these measures, multiple prediction equations were obtained using the 18 CPI scales and the GT measure as the independent variables with academic grades as the dependent variable. The results of the Wherry-Doolittle variable selection procedure indicated that the following combinations of variables had the maximal correlation with the criterion: NP class—GT + Ai and GT + Ai + Sa; CP + SW classes—GT + Ai and GT + Ai + Cm. The resultant multiple prediction equations were:

NP:

$$\hat{X}_0 = .293 + .319(GT) + .589(Ai)$$
$$\hat{X}_0 = .176 + .316(GT) + .658(Ai) + .497(Sa)$$

CP+SW:

$$\hat{X}_0 = .374 + .306(GT) + .387(Ai)$$
$$\hat{X}_0 = .172 + .284(GT) + .439(Ai) + .845(Cm)$$

where the formulas are in raw-score form and $\hat{X}_0$ = predicted criterion score.

Hence, there were two equations per group of *Ss*, a two-variable and a three-variable, as they will be called. The next step in the study was concerned with a cross-validation of these equations on three new classes. Only two- and three-variable equations were considered, since the addition of a fourth variable produced a negligible increase in the multiple R in the original sample. The pre-

[5] The GT score is an average of the scores obtained on the Reading and Vocabulary (RV) and the Arithmetic Reasoning (AR) subtests of the Army Classification Battery. Hedlund (1957) reports that a GT range of 90–109 approximates a Wechsler-Bellevue IQ range of 94–107 and a Wechsler Adult Intelligence Scale IQ range of 92–106. The RV and AR subtests have also been found to correlate well with the Wechsler-Bellevue Form 1 (Montague, Williams, Lubin, & Gieseking, 1957).

TABLE 3

CROSS-VALIDATED VALIDITIES FOR THE OBTAINED MULTIPLE PREDICTION EQUATIONS

Prediction equations	Multiple *R*s of original sample		Cross-validation *r*'s of new sample	
	CP + SW ($N = 98$)	NP ($N = 64$)	CP + SW ($N = 35$)	NP ($N = 33$)
GT + Ai	.610**	.600**	.631**	.558**
GT + Ai + Cm	.639**	—	.566**	—
GT + Ai + Sa	—	.641**	—	.558**

** $p < .01$.

dicted criterion scores were obtained for each *S* by solving the above equations. Correlations were then obtained between the predicted and obtained criterion scores for each class. The results are presented in Table 3.

DISCUSSION

The correlations presented in Table 2 indicated that the CPI scales are measuring personality attributes which are related to academic success. In terms of the single measures presented in Table 2, however, it was also obvious that the GT variable, a purely intellectual measure, is superior to any of the CPI scales in predicting academic achievement. On the other hand, a combination of a measure of intellectual ability (GT) and a measure of personality functioning (the Ai scale of the CPI) is superior to any one of these variables taken alone.

The Ai scale (Achievement via Independence) is described by Gough as having the following purpose: "To identify those factors of interest and motivation which facilitate achievement in any setting where autonomy and independence are positive behaviors." High scores are described (Gough, 1960) as "Mature, forceful, strong, dominant, demanding, and foresighted; as being independent and self-reliant; and as having superior intellectual ability and judgment" (p. 13).

It appears, then, that this combination of variables will enable one to assess both the intellectual and motivational factors that determine the degree of success that a student will achieve in these particular areas of training. The degree to which this measure is able to predict is quite surprising due to the fact that we are dealing with a fairly "select" group. The students in these classes have already successfully completed basic military training (8 weeks) and an introduction to military medical activities (2 weeks for the CP and SW classes and 8 weeks for the NP class). There has already been much "weeding out" of those who have difficulty in adjusting to the demands of military life.

The data also indicate that the CPI is a much more useful instrument for differentiating poor and good students in the CP, SW, and NP classes than was the MMPI. Hewitt and Rosenberg (1961) found that successful graduates in these three courses had MMPI profiles that were more abnormal than those of students who failed the courses.

REFERENCES

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1960.

HEDLUND, J. L. Miscellaneous notes regarding army classification scores and general intelligence. Washington, D. C.: Walter Reed Army Hospital, 1957. (Mimeo)

HEWITT, J. H., & ROSENBERG, L. A. The MMPI as a screening device in an academic setting. *Educ. psychol. Measmt.*, 1962, **22**, 129–137.

MONTAGUE, E. K., WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. F. The use of army tests for the assessment of intellectual deficit. Research Report No. 52-57, 1957, Walter Reed Army Institute of Research, Washington, D. C.

(Received July 20, 1961)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 389–392

EFFECTS OF BRANCHING IN A COMPUTER CONTROLLED AUTOINSTRUCTIONAL DEVICE

JOHN E. COULSON, DONALD P. ESTAVAN, RALPH J. MELARAGNO, AND HARRY F. SILBERMAN

System Development Corporation, Santa Monica, California

2 groups of 15 high school Ss received instruction in logic from a computer controlled autoinstructional device. In 1 group all Ss received a fixed sequence of 233 items. In the 2nd group each of the Ss received a different number and sequence of items, depending on the S's performance during the lesson. Branching decisions were based on errors and on the S's evaluation of his own readiness to advance to new topics. Posttest scores were significantly higher (.05 level) for the branching group than for the fixed sequence group; training time differences were not significant.

Experiments comparing fixed-sequence with variable sequence methods of autoinstruction have yielded equivocal results. One study, using manually controlled autoinstructional devices and a highly simplified form of branching based on student errors, showed a saving in training time for the variable sequence method but no difference in posttraining criterion performance between this mode and the fixed sequence mode (Coulson & Silberman, 1960). In a second experiment (Silberman, Melaragno, Coulson, & Estavan, in press), a computer controlled autoinstructional device provided a branching pattern with more alternative paths, but with student errors in training items again used as the sole branching criterion. This method was found not to differ significantly in either training time or criterion performance from a fixed sequence procedure.

Another study, using manual presentation techniques, compared a fixed sequence mode with two variable sequence procedures in which the student was allowed to branch at his own option (Silberman, Melaragno, Coulson, & Estavan, in press). One of the two variable sequence procedures was found superior to the fixed sequence method on criterion performance, with no difference in training time.

One factor apparently influencing the success of variable sequence autoinstruction is the nature of the branching criteria used to determine what item sequence should be given to each student. Such criteria should provide accurate evaluations during the lesson of the student's knowledge of topics which he is expected to learn. From these evaluations it should be possible to determine whether the student is ready to proceed to a new topic, or whether he requires additional instruction on the same topic.

The present experiment was designed to explore the effectiveness of a procedure in which both the student's errors on training items and his evaluation of his own learning progress were used as branching criteria. It was hypothesized that a combination of these two indices would provide effective evaluation of student progress, and that students taught with this procedure would score higher on a posttraining criterion test than another group receiving fixed sequence autoinstruction. It was further predicted that the variable sequence group would require less training time than the fixed sequence group.

METHOD

Two groups of 15 high school students were given automated instruction on logic by means of a flexible sequence, computer controlled autoinstructional device. One group was designated the "fixed sequence group," and the other was termed the "branching group." Every member of the fixed sequence group received the identical sequence of 233 instructional items. Each student in the branching group was given a different number and sequence of items, depending on his performance in responding to certain items during the lesson and on the student's self-evaluation of his own progress.

A brief rest period was given every student toward the middle of the lesson. At the conclusion of the lesson each student received a written criterion test covering the materials in the lesson. This test was the same for all members of both groups.

Equipment

The autoinstructional device used in this study has been described in an earlier report (Coulson & Silberman, 1961). It consisted of a general purpose digital computer (Bendix G-15) coupled to a random-access slide projector and an electric typewriter. The slide projector acted under computer control in selecting a sequence of slides, each of which contained one instructional item. The subject used the typewriter to insert his answers to multiple-choice questions contained in the items, and to receive knowledge of results concerning the accuracy of his answers.

For the branching group the computer was programed to select item sequences on the basis of three student response criteria: (*a*) number of response errors cumulated over a set of items relating to the same topic, (*b*) answers given by the subject to individual questions where the erroneous alternative selected is indicative of a particular misunderstanding by the subject, (*c*) answers given by the subject to items asking whether he feels ready to advance to a new topic or is in need of further review on an earlier topic.

If a member of the branching group made too many errors on a particular topic or expressed a lack of confidence in his own progress, he was given extra remedial materials to strengthen his knowledge of that topic. Otherwise he was taken immediately to a new topic.

For the fixed sequence group the device merely presented each of the 233 items in sequence.

Instructional Materials

The 345 instructional items on logic used in this study were revised from materials used previously by Silberman, Melaragno, Coulson, and Estavan. Topics covered by these items included simple and compound statements, logical connectives, and six argument forms.

Materials were organized into "main stream" items using fairly large and difficult steps to cover topics, and remedial items covering the same topics with smaller and more numerous steps. Some items were written as statements, while other items were written as diagnostic test questions. The diagnostic items, which formed the major basis for branching decisions, usually included several plausible distractors and were designed to be highly discriminating of student knowledge.

A number of self-evaluation items was included in the revised materials. These items asked the student to judge his own state of knowledge on different topics.

The 233 items given to all members of the fixed sequence group were selected by an independent consultant who was not associated with the experiment, nor with the original item preparation. The consultant was experienced in the development and experimental evaluation of self-instructional materials and familiar with the subject area being taught. He was given the full set of 345 logic items and a list of the topics covered by the criterion test. He was asked to select from the items a sequence which he felt would provide effective instruction on the topics covered by the criterion test, and was also allowed to add any items which he believed appropriate.

Subjects

Paid volunteers from the junior and senior classes of two high schools in the Los Angeles area served as subjects in the experiment. Students were assigned to two groups of 15 subjects each. Each group contained students from both schools. The two groups were matched for average scores on the Henmon-Nelson Test of Mental Ability (Form A, Grades 9–12) administered before training. Henmon-Nelson means for the fixed sequence group and the branching group were 112.5 ($SD = 12.9$) and 113.8 ($SD = 15.9$), respectively. Subjects were naive with respect to the subject matter (logic) used in the lesson.

Criterion Test

The posttraining criterion consisted of 51 multiple-choice and 44 fill-in test items. Some of the questions were quite similar to the training materials while others required transfer to situations not encountered by the subject during training. The split-half reliability of the test was .97 (after Spearman-Brown correction).

RESULTS

The mean number of instructional items seen by members of the branching group was 213.5 with a standard deviation of 19.4. All members of the fixed sequence group received 233 items.

Required training times and posttraining criterion scores for the two groups are summarized in Table 1.

TABLE 1

TRAINING TIME AND POSTTRAINING CRITERION PERFORMANCE FOR EXPERIMENTAL GROUPS

	Treatment group	
	Branching[a] N = 15	Fixed sequence[b] N = 15
Training time (minutes)		
M	116.2	129.5
SD	20.7	26.8
Posttraining criterion test scores		
M	74.3	65.4
SD	10.7	14.9

[a] Average 213.5 item.
[b] 233 items.

Members of the branching group averaged 8.9 points higher on the criterion test and required an average of 13.3 minutes less training time. The difference in criterion performance was significant at the .05 level ($t = 2.15$, $df = 28$). A one-tailed test was used since the direction of the difference was predicted in advance. The difference in training time did not reach significance.

The correlation between aptitude score and posttraining criterion score was not significant for either group nor for the two groups combined.

DISCUSSION

Results of the experiment indicate an advantage in favor of the branching procedure over the fixed sequence mode of instruction. Student training errors and responses to self-evaluation questions provided effective criteria for branching, presumably by serving as accurate indices of student learning progress.

In an earlier study, using the same autoinstructional device and covering the same topics in logic, no difference was found between a branching and a fixed sequence mode of instruction (Silberman, Melaragno, Coulson, & Estavan, in press). A comparison of materials and procedures used in the two experiments reveals several dissimilarities which may account for differences in the results:

1. Branching in the first experiment was based purely on training errors. The use of self-evaluation items as additional cues for branching in the present study probably increased the accuracy of evaluation of student progress.

2. In the first study practically all items contained questions. Thus most items were used for both instruction and diagnosis. In the present study some items contained only statements, while other items were written specifically as diagnostic test questions. It is probable that the use of specially designed diagnostic items afforded better estimates of student readiness to advance to new topics in the lesson, and thus made the branching more appropriate to individual student needs.

3. A remedial sequence in the earlier study contained more items with smaller steps than the main stream segment which it was intended to supplement, but both main stream and remedial items used the same basic approach in presenting the facts or concepts to be learned. The revised remedial items used in the present experiment approached topics differently than their corresponding main stream items. Thus, where the main stream items taught a particular topic deductively the remedial items might teach the same topic inductively. It is likely that the revised remedial materials were more effective in allowing students who had encountered obstacles on the main stream items to view the concepts being presented in a new perspective.

4. In the first study each member of the fixed sequence group received the same sequence of items as a "mate" in the branching group. The only difference between the two groups was that the sequence was determined on the basis of errors made during the lesson by the branching group member, whereas this same sequence was given to the fixed sequence group member without regard to his performance on the materials. To further equate conditions for the two groups, knowledge-of-results messages for both groups contained only the number of the correct answer. One consequence of the attempt to hold conditions identical for the two groups in the first experiment was that certain types of branching and certain remedial materials could not be used in the branching condition. For example, no remedial items or knowledge-of-results messages could be given that referred specifically to the student's answer to the previous item, since such a message would have been inappropriate for the fixed sequence group mate who might have given a different answer. In the present experiment no attempt was made to equate exactly the materials seen by the two experimental groups. Members of the branching group were often given remedial items containing specific references to answers which they had given to preceding items. These remedial materials were designed to correct misunderstandings evidenced in the earlier responses. A remedial item, for example, might say, "No, you gave the answer '$\vee$' when you should have answered '$\wedge$.' Remember, '$\wedge$' represents 'and'

while 'V' represents 'or.'" Thus the revised materials used in the present experiment were much more closely adapted to the needs of individual branching group members than was possible in the initial study.

In summary it appears that the branching procedure used in the present experiment provided more accurate diagnosis of student needs than was possible in the earlier study, and more effective remedial materials for filling these needs. Some support for this conclusion is given by the fact that the correlation between student aptitude and criterion score was significant in the first study, but not in the second. This finding indicates that the remedial materials in the later experiment raised the performance of lower aptitude students to that of the higher aptitude students, although the lower aptitude students required more training time to reach this level. Since there was no aptitude-criterion correlation in the second study for either the branching group or the fixed sequence group the reduction of student performance differences cannot be attributed entirely to the branching procedure. The fact that some of the items used for remedial purposes in the branching group were included in the materials selected for the fixed sequence group may account for the lack of correlation between aptitude and criterion score in the fixed sequence group.

REFERENCES

COULSON, J. E., & SILBERMAN, H. F. Effects of three variables in a teaching machine. *J. educ. Psychol.*, 1960, **51**, 135–143.

COULSON, J. E., & SILBERMAN, H. F. Automated teaching and individual differences. *Audiovis. commun. Rev.*, 1961, **9**, 5–15.

SILBERMAN, H. F., MELARAGNO, R. J., COULSON, J. E., & ESTAVAN, D. P. Fixed-sequence versus branching auto-instructional methods. *J. educ. Psychol.*, in press.

(Received September 5, 1961)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 393–398

DECEPTIVE PACKAGING:
ARE THE DECEIVERS BEING DECEIVED?

JAMES C. NAYLOR[1]

Ohio State University

To investigate the effects of deceptive packaging, sample packages of potato chips containing various amounts were distributed to consumers who purchased a regular pack of the same brand. The only difference between the sample and the purchased pack was in terms of quantity of chips with the sample packs having various deficits. Preference data were obtained via a follow-up interview. The analyses of the questionnaire data indicated that: (a) consumers did not perceive weight differences between packs, (b) all indices of consumer preference favored the free pack when the two were of equal weight, and (c) all indices showed a marked decrease in preference for the experimental pack as the contents of the experimental packs were decreased.

The tremendous emphasis placed upon packaging as a wedge to success in the consumer market in recent years has resulted in certain packaging techniques which have been questioned on an ethical basis. The most outstanding of these is the practice known as "deceptive packaging." Briefly defined, deceptive packaging involves a deliberate attempt to mislead the consumer regarding some aspect of the product (usually quantity) by using packages specifically designed to convey a false impression concerning this product attribute (e.g., making the package oversize). The extent to which this practice has become prevalent is difficult to state, although the fact that in recent months a Senate subcommittee was formed to inquire solely into this particular problem would attest to its frequency of occurrence.

From the point of view of consumer psychology, the problem is not primarily one of ethics, but rather of determining what effect deceptive packaging has upon the consumer, per se. First, does he realize that he is being deceived? Second, what is the effect of the deception upon his perception of the product (*a*) if he feels he has been deceived, or (*b*) if he is not aware of any deception? The purpose of this study was to experimentally investigate a case of deliberate deceptive packing in an attempt to provide answers to some of these questions.

[1] The author wishes to acknowledge the data collection efforts of G. Croft Henry, John Mitchell, and Don Shumaker, members of the research field team.

Method

The independent variable chosen for study was package quantity. Through the cooperation of a local potato chip manufacturer, it was possible to obtain 144 experimental twin packs which were filled to specified weight requirements. The normal contents of a regular twin pack in terms of weight was 9 ounces. The experimental twin packs were of three weights: 9 ounces (control packs), 8 ounces, and 7 ounces. The experimental packs were all filled with chips from the same batch in order to assure equal treatment. All were packaged in regular twin pack containers in a normal manner. The only difference between the experimental packs was in terms of quantity as determined by weight. Two retail outlets were selected for distribution: a medium sized independent grocery located in a suburban area (Store A) and a large chain store market (Store B) in more metropolitan surroundings.

Procedure. Because of possible ill feeling which might have been generated toward the donor company, it was not feasible to have the experimental packs actually purchased by the consumer. Thus, a procedure was adopted in which the experimental packs were given, free of charge, to any person willing to participate in the experiment. An observer was stationed with a supply of the 7-, 8-, and 9-ounce experimental packs within view of the normal store display. Any consumers who "purchased" (i.e., clearly selected from the display) a twin pack of the donor brand were approached by the observer and asked as to their willingness to cooperate in a university study of consumer preferences. They were told that they would receive a sample pack free of charge on the condition that they were to "compare the contents of the two packs (regular versus experimental) in terms of personal preference." This preference would be obtained by a follow-up 10-minute interview in their own home within 3 days.

If the customer was agreeable to this, he or she was then given one of the experimental packs as a gift (the three weights were distributed in random

order) and a time and day for the subsequent interview were agreed upon.

A reminder slip with the time and date for the interview was made out and stapled to the sample (free) pack. This slip also served to provide the basis for differentiation between the sample and the regular pack.

Distribution of packs was done only between the hours of 4:00 P.M. and 9:00 P.M. on Thursdays, Fridays, and Saturdays. Twenty-four experimental packs were distributed every week—eight of each weight. This procedure was continued for a 6-week period. To prevent staleness of the experimental chips, they were obtained from the donor once every 2 weeks in lots of 48 (16 of each weight). Thus, over the 6-week period a total of 144 experimental packs was distributed—48 of each weight class. The first 3 weeks of distribution was in Store A, the remainder in Store B.

Interview. The interview itself consisted of having the consumer fill out a short, nine-item questionnaire (see Table 1). The first five items were for the purpose of obtaining information desired by the donor company, and Questions 6 through 9 were

TABLE 1

SAMPLE QUESTIONNAIRE USED IN THE INTERVIEW

AGE_____

SAMPLE_____

SEX_____

1. Was this the first time you have purchased this brand? Yes___ No___
2. Do you usually purchase this brand? Yes___ No___
3. Do you expect to purchase this brand again? Yes___ No___
4. What did you like most about this brand?_____
5. What (if anything) did you like least about this brand?_____
6. Please circle the point on the scale that expresses your reaction to each pack.
   a. How well did you like the contents of the regular pack?

1	2	3	4	5	6	7	8	9
terrible		poor		average		good		excellent

   b. How well did you like the contents of the sample pack?

1	2	3	4	5	6	7	8	9
terrible		poor		average		good		excellent

7. If both were available for purchase the next time you went to the store, which would you buy?
   ___ sample pack
   ___ regular pack
   ___ either
   ___ neither
8. List three words that describe the two packs.
   a. Sample pack: _____
   b. Regular pack: _____
9. Which pack had the most in it?
   ___ regular pack
   ___ sample pack
   ___ same

designed to obtain preference information as a function of perceived deception. Any requests for additional insight into the experiment on the part of the consumer at the time of the interview were refused. However, a promise to mail a short explanation at the conclusion of all interviews was made where desired. All such requests were subsequently honored.

Interviewers. The distribution of packs and the subsequent interviewing of consumers were done by a three-man research team consisting of a senior psychology graduate student and two undergraduate psychology majors.

RESULTS

Sample Characteristics. Only six customers declined to participate in the survey when approached by the observer, making the total acceptances 144 out of 150. This represents a loss of only 4%. Out of the 144 customers who agreed to participate, however, 26 were lost during the interview phase—primarily due to absences as a function of summer vacation. Thus, the final data sample consisted of 118 consumers. Table 2 shows the age and sex characteristics for the 114 people who responded to these two interview questions.

As indicated in the table, the sample was predominantly female. A chi square test was performed on the sample frequencies to determine if the relative proportion of males and females differed for the three experimental groups. The results clearly indicated that they did not ($\chi^2 = .167$, $df = 2$, $p > .99$). In addition, a simple analysis of variance test was made to test for age differences. The mean ages for the three experimental groups were not significantly different ($F = .04$, $df = 2/113$).

In order to discuss the results of deceptive packaging, it was necessary to have some indication of the extent to which the weight shortages in the sample packs were perceived by the consumers. Table 3 shows the frequencies of the responses to the categories of Question 9 on the questionnaire. Regardless of experimental pack weight, the majority of the responders indicated the two packs as being equal in contents. A chi square test of significance showed no significant shift in proportions as the contents of the experimental pack was decreased ($\chi^2 = 3.121$, $df = 4$, $p < .50$).

It was possible to obtain two judgment scores for each person—his judgment for the regular pack and his judgment for the experimental pack—from Question 6. A difference score was then computed for each individual using the following formula: Experimental Pack Scale Value − Regular Pack Scale Value.

This resulted in a score indicating the degree to which the contents of the experi-

TABLE 2

BREAKDOWN OF CONSUMER SAMPLE IN TERMS OF AGE AND SEX

Criterion	Experimental group			Total
	9 ounces	8 ounces	7 ounces	
Age *M*	36.7	37.5	38.7	37.6
Age *SD*	10.7	12.3	7.7	10.2
Male	6	4	5	15
Female	35	31	33	99
N	41	35	38	114

TABLE 3

FREQUENCY OF PEOPLE JUDGING EXPERIMENTAL OR REGULAR PACK AS HEAVIEST AS A FUNCTION OF EXPERIMENTAL PACK WEIGHT

($N = 118$)

Weight of experimental pack	Pack judged heaviest		
	Experimental	Regular	Equal
9 ounces	9	4	30
8 ounces	5	8	22
7 ounces	7	8	25

Note.—$\chi^2 = 3.121$, $df = 4$, $p < .50$.

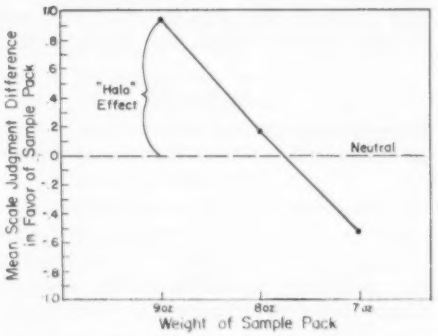

FIG. 1. Preference in favor of the contents of the experimental pack as a function of experimental pack weight.

TABLE 4

ANALYSIS OF VARIANCE OF THE DIFFERENCES IN THURSTONE JUDGMENTS OF PREFERENCE BETWEEN REGULAR AND EXPERIMENTAL PACKS AS A FUNCTION OF EXPERIMENTAL PACK WEIGHT

($N = 118$)

Source	*df*	*MS*	*F*
Weight	2	22.0	4.58*
Error	115	4.8	
Total	117		

* $p < .05$.

mental pack were preferred over those of the regular pack. The mean of these scores was then computed for each group (see Figure 1). When the two packs were of equal weight a marked preference was indicated for the experimental pack. When the experimental pack was 1 ounce less, however, the preference for it was only slight, and when there was a 2-ounce difference the regular pack was substantially preferred. An analysis of variance of these judgment differences (see Table 4) indicated that the decrease in preference for the experimental pack as a function of its weight was statistically significant at the .05 level ($F = 4.58$, $df = 2/115$).

A second measure of the effect of weight on preference was obtained by analyzing the responses to Question 7. Table 5 shows the distribution of responses to the alternatives presented by this question. The response alternative "neither" only attracted a total of four responses and was therefore eliminated in the analysis. When the packs were of

TABLE 5

DISTRIBUTION OF PREDICTED NEXT PURCHASE CHOICE AS A FUNCTION OF EXPERIMENTAL PACK WEIGHT

($N = 118$)

Weight of sample	Predicted choice		
	Buy experimental	Buy either	Buy regular
9 ounces	22	7	11
8 ounces	13	14	6
7 ounces	12	8	15

Note.—$\chi^2 = 10.203$, $df = 4$, $p < .05$.

TABLE 6

SCALE VALUES OF DESCRIPTIVE ADJECTIVES USED TO DESCRIBE SAMPLE AND REGULAR PACKS

Mean scale value	Words or phrases
4.6	Crispy, Delicious
4.5	Very tasty, Crisper
4.4	Very very good, Very fresh, Excellent, Superb
4.3	Scrumptious, Very good, Fresh
4.2	Tastier, Crisp (as usual), Tasty
4.1	Always good
4.0	Good taste, More crisp, Crisp, Full of goodness
3.9	Good color, Flavorful, Better flavor, Appetizing
3.8	Lighter, Good, Flavorsome, Good smell, Better taste
3.7	Better, Clean taste, Crunchier
3.6	Wholesome, Satisfying, Good size, Richer, Crisp edges
3.5	(None)
3.4	Good oil, Not greasy, Tangy, Not too salty
3.3	Well done, Large, Thinner, Uniform size, Thin, Greaseless, Smooth texture
3.2	Digestible, Lighter color, Bigger
3.1	Done, Less oily, Dryer
3.0	Brown, Not burnt, Palatable, Salty, Smooth
2.9	More tender, Saltier, Shorter, Same
2.8	Not too oily, Brown edges
2.7	Milder, Small, Not as large
2.6	Not as thin, Unsized, Greasy but good, Less salty
2.5	Stronger taste, Average, Rough, Heavier, Too thin
2.4	Darker
2.3	Too well done, Ordinary, Thicker
2.2	Not tender, Cheesy taste, Too brown
2.1	Too small, Not enough shortening, Too salty
2.0	Less flavorful, Heavy texture, Not very good, Not as crisp
1.9	Oily, Burnt, Green, Scanty amount
1.8	Hard, Not crisp, Unseasoned
1.7	Flat
1.6	A little greasy, Mealy
1.5	Greasy, Undesirable
1.4	Slightly stale, Greasy taste
1.3	Bitter, Limp, Tasteless, Too greasy

equal weight, the majority of the consumers stated that, if given a choice, they would purchase the experimental chips. With a 1-ounce weight differential the plurality response changed to "would buy either," and with a 2-ounce difference the plurality response was in favor of the regular pack. A chi square test indicated that this change in preference was significant at the .05 level ($\chi^2 = 10.203$, $df = 4$).

A third analysis of package preference was obtained through the descriptive adjectives applied to each pack in Question 8. A total of 109 different adjectives or adjective phrases were used by the consumers in describing the two packs. These phrases were given to 10 judges (psychology graduate students) who rated each of them for favorability on a five-point Thurstone rating scale. The phrases and their resulting mean scale values are shown in Table 6. The analysis of variance reliability for the ratings was .96, and the difference between rater zero points was significant at the .01 level ($F = 23.95$, $df = 108/792$).

Each person was then assigned two mean scores on the basis of the adjectives used to describe the regular and experimental packs using the favorability scale value for each adjective. The difference between these two means was then computed as follows: ($\bar{X}$ adjective scale value for experimental pack) − ($\bar{X}$ adjective scale value for regular pack). This resulted in an adjective preference score for the experimental pack for each person. The mean of these adjective preference scores was then computed for each group (see Figure 2). The same pattern of preference as a function of experimental pack weight was found with the adjective values as had been found previously with the Thurstone judgments. As the weight of the experimental pack contents was decreased, the preference for that pack also decreased. The analysis of variance on the means in Figure 2 indicated that they were not significantly different from one another ($F = .34$, $df = 2/115$) even though the differences were in the expected direction.

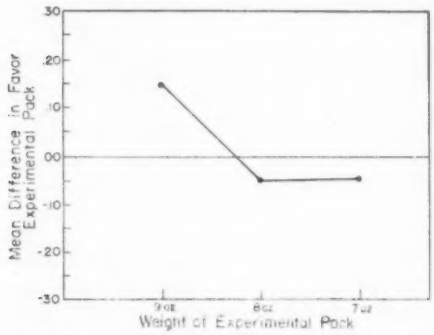

FIG. 2. Mean difference in scale values of adjectives used to describe experimental and regular packs as a function of experimental pack weight.

DISCUSSION

Deception in packaging can occur in two general ways. First, a consumer may be deceived at time of purchase and subsequently discover this deception at the time of product consumption. Such a case may be considered a "purchase deception." Second, a consumer may be deceived at time of purchase and not discover the deception at all. The latter may be considered a "consumption deception." Since the data from Question 9 indicated that even following the actual consuming of the chips people were not aware of any shortages, the results of the study may be considered as applying to a consumption deception and not to a purchase deception.

From an ethical point of view the consumption deception is the most severe, since the consumer, unaware of anything amiss, will supposedly continue to be satisfied with his purchase. With a purchase deception, however, he is overtly aware of having been taken advantage of and is thus able to modify his subsequent buying behavior to avoid a repeat deception. The results of the several preference indices computed in this study, however, indicate that even in the case of a consumption deception there is created a dissatisfaction with the product, even though the source of dissatisfaction cannot be identified by the consumer! As the contents of the experimental pack were decreased, there was a related decrease in product preference for the experimental pack. This occurred for all three indices—significantly so for two of them. The non-

significance in the third case may have been a function of limited variance due to the use of a five-point Thurstone scale in scaling the adjectives.

The fact that the experimental pack was markedly preferred with all indices when it was the same weight as the regular pack was probably a function of (*a*) the fact that it was free, and (*b*) an aura of "if its something new, it must be better" surrounding the free pack. The end result was to produce a halo effect (see Figure 1) which accounted for the experimental pack preference.

If indeed the answers to Question 7 were true predictors of future buying decisions, the fact that a consumption deception had occurred would result in a decrease in purchases for the pack having a weight shortage, and this decrease would be related to the degree of deception involved. While such extrapolation of buying behavior on the basis of questionnaire responses is sometimes dangerous, the implications should be obvious to those engaged or planning to engage in a program of package deception.

(Received October 31, 1961)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 399–400

CLASS ASSIGNMENT IN THE LATENT PROFILE MODEL

W. A. GIBSON

Personnel Research Office, Department of the Army, Washington, D. C.

The latent profile model provides a way of establishing an empirically-based typology from quantitative measures, and this paper sketches a method for assigning any person's score profile to its most appropriate type.

Latent profile analysis has been described elsewhere (Gibson, 1959) as the generalization to quantitative data of latent structure analysis (Lazarsfeld, 1959). The latent profile analysis of intercorrelations r_{jk} among n continuous variables divides the total sample into a set of q subsamples or *latent classes*, for each of which is given its size and its profile of standard score means on the n variables. These *latent profiles* are then to be compared and interpreted as to the kinds of people they characterize. Thus latent profile analysis provides a way of evolving empirically-based typologies such as personality categorization, psychiatric syndromes, job interest constellations, or modes of political behavior.

The latent structure and latent profile models attribute all observable covariation among the n variables to variation *between* the latent classes, so that within-class variation is, by definition, statistically independent. This is the *local independence* definition of a latent class. The rationale for this definition is that if a class is made homogeneous enough, with respect to whatever and however many underlying dimensions or "factors" may be involved, then all within-class covariation attributable to those underlying dimensions will have been partialed out.

Given the results of a latent profile analysis, how can subjects be assigned to their most appropriate latent class? For this purpose consider the n-dimensional scatter diagram of standard scores, and, for concreteness, imagine $n = 3$. Imagine, further, that two latent classes suffice to account for the intercorrelations, or that $q = 2$. Now each latent profile defines a point in that three-space, and the members of the associated latent class are presumably, by virtue of local independence, clustered randomly around that point. Thus an equal-density contour of the class is an ellipsoid with its three axes parallel to those of the scatter diagram as a whole. The comparable contour of equal relative density for the other class is, assuming homoscedasticity of within-class variance, an ellipsoid of identical size and orientation, but with different position.

The length of each semiaxis of such an ellipsoid is determined by the within-class variance of the associated variable. Since we are dealing in standard score units, the within-class variance of Variable j is the complement of (or one minus) the between-class variance, r_{jj}, obtainable from the latent profile solution. Then the length of Semiaxis j of the ellipsoid that corresponds to one unit of within-class variance is the square root of the within-class variance of Variable j, or is simply $\sqrt{1 - r_{jj}}$.

The n-dimensional scatter diagram containing q families of concentric ellipsoids may be converted into a system of spheres by the simple trick of dividing all standard scores on each Variable j, including those in the latent profiles, by $\sqrt{1 - r_{jj}}$. This merely stretches the various axes so that spheres result, and it makes the unit of measurement in all directions be the square root of within-class variance. Now each person might possibly be assigned to that class to whose center he is closest. Squared distances, computed either as sums of squared differences between coordinates or by the cosine law in the "spherized" scatter diagram, would suffice for that purpose. With equal class sizes, such a procedure would amount to assigning each person to the latent class from which he is most likely to have come. However, varying class size must enter the picture since, for example, a person who is equidistant from the centers of two classes of unequal size is more likely to have come from the larger.

With unequal class sizes, a maximum likelihood assignment procedure is a little more elusive, though not so much as might first

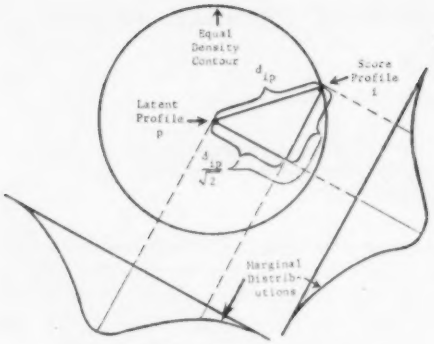

FIG. 1. Latent class p for $n = 2$.

appear. Local independence and the spherized space make the relative frequency of Score Profile i in Latent Class p depend only upon the associated distance, d_{ip}, and the dimensionality, n, of the space. Figure 1 illustrates why this is so for the case where $n = 2$. No matter what direction Point i is from the center of the class, p, an isosceles right triangle can be constructed as shown with squared side equal to $d_{ip}^2/2$ by virtue of the Pythagorean theorem. Now imagine a bivariate normal surface with $\rho = 0$ centered upon Point p. Its height at Point i is, because $\rho = 0$, simply the product of associated marginal frequencies. These two marginal frequencies become identical when the axes are placed parallel to the legs of the isosceles right triangle, and they are then proportional to the ordinate of a unit normal curve at Point $d_{ip}/\sqrt{2}$. This line of argument generalizes to n space, wherein the relative frequency of Profile i in Class p is proportional to the nth power of the unit normal ordinate at Point $d_{ip}/\sqrt{n}$. Expressed in terms of the proportionality constant c and the equation of the unit normal curve, the marginal frequency in question becomes

$$\frac{c}{\sqrt{2\pi}} \cdot e^{-(d_{ip}^2/2n)}, \qquad [1]$$

and its nth power is

$$\left(\frac{c}{\sqrt{2\pi}}\right)^n \cdot e^{-(d_{ip}^2/2)}. \qquad [2]$$

Multiplying the latter by the Class Size f_p gives

$$\left(\frac{c}{\sqrt{2\pi}}\right)^n \cdot f_p \cdot e^{-(d_{ip}^2/2)} \qquad [3]$$

as the absolute number of people with Profile i who belong to Class p. Computation of this absolute frequency across all of the latent classes provides the means for assigning all people with Profile i to the class to which the most such people belong. The amount of misclassification will then be the sum of the frequencies for that profile in the remaining classes.

The natural logarithm of Expression 3 is

$$n \ln\left(\frac{c}{\sqrt{2\pi}}\right) + \ln f_p - \frac{d_{ip}^2}{2}, \qquad [4]$$

in which the first term is constant and the second depends only on class size. Dropping the first term and doubling the last two gives

$$2 \ln f_p - d_{ip}^2, \qquad [5]$$

an expression monotonically related to Expression 3, hence just as good for assignment purposes. Expression 5 shows that distance alone suffices for maximum likelihood assignment with equal class sizes, and that class size alone determines assignment for equal distances.

Assignment to type is an important aspect of a typology that has already been established, but such assignment can also lend concreteness in the early stages of establishment, where having actual examples of the types helps to suggest what kinds of people they are.

REFERENCES

GIBSON, W. A. Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 1959, **24**, 229–252.

LAZARSFELD, P. F. Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. *Formulations of the person and the social context*. New York: McGraw-Hill, 1959.

(Received November 13, 1961)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 401-408

PRODUCTIVITY AND ERRORS IN TWO KEYING TASKS: A FIELD STUDY

E. T. KLEMMER [1] AND G. R. LOCKHEAD [2]

International Business Machines Research Center, Yorktown Heights, New York

Productivity and error rates were measured for a billion responses by more than a thousand operators of IBM card punches and bank proof machines in 20 different installations. Productivity increases and errors decrease during the 1st year on the job, sometimes longer. Experienced card punch operators average 56,000 to 83,000 keystrokes per day with 1600 to 4300 strokes per undetected error. Experienced bank proof machine operators average 4350 to 6600 checks per day with about 3500 checks per undetected error. The fastest operators at any installation produce twice as much as the slowest. The least accurate operators make 10 times as many errors as the most accurate. Fast operators tend to make fewer errors ($r = -.5$).

A conservative estimate is that two million people spend a large part of each day before a keyboard driven machine: typewriter, keypunch, adding machine, teletype, bank proof machine, linotype, etc. The productivity and error rates of these operators are of interest to the companies that employ them, the companies that design and build the machines, and the psychologists who are interested in this form of human behavior.

There is a long history of studies concerning keyboards and keying, mostly centered around the typewriter. Many studies looked at training (e.g., West, 1957) and several comprehensive publications provide rules for setting work standards for keyboard driven machines based on time and motion analyses (e.g., Colver, 1960). Studies have been aimed at designing new and better keyboards (e.g., Hillix & Coburn, 1961), and several papers review the work done thus far with typewriters (e.g., Rahe, 1954).

The present study does not add laboratory or theoretical data to this field but rather presents data on production and error rates from a large number of operators in the course of their actual daily work. These data will be useful in providing a baseline for experimental work as well as describing the distribution of production and error rates for keyboard operators.

The data in this report were drawn from more than a billion responses made by more than a thousand operators in 20 installations. The experience of the operators ranged from 1 week to several years. All of the data presented are from two types of machines: the IBM card punch Models 024 and 026, and the IBM bank proof machines, Models 802 and 803.

Most of the current report is based on records routinely kept by large installations of card punches and bank proof machines. The operators were not told that their performance records would be subjected to special analysis. In addition, the results of two special speed tests are given for comparative purposes.

Few installations keep uniform records of the errors made by operators in addition to their production records. Since error rates are less easy for the research worker to determine than are production rates, considerable attention has been given to errors. Fortunately, two large installations did keep regular and uniform records of both production and error rates for all operators.[3] The error analyses presented are based largely on the data from these: one keypunch and one bank proof machine installation.

Production and error figures are treated separately with each broken down according to differences occuring between installations, differences between operators in the same

[1] Now at Bell Telephone Laboratories.

[2] Now at Johns Hopkins University.

[3] We are particularly indebted to C. R. Pricher of the New York Federal Reserve Bank, and S. Kreps and J. Roarty of the Social Security Administration for their cooperation in making comprehensive records available for this study.

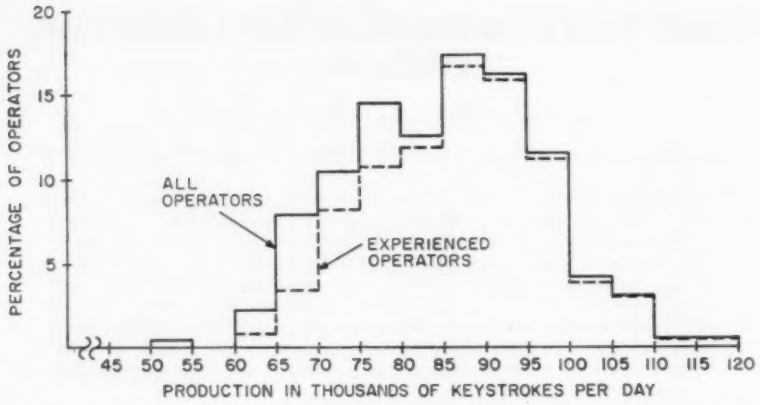

FIG. 1. Distribution of 271 card punch operators according to average daily production rate.

installation, and variability within operators. Many reasons could be cited to explain these differences but most would be speculative. We shall refrain from such speculation as much as possible.

Some interesting consistencies in error rates are shown which immediately raise the hope of a simple model to describe at least some of human error making.

DESCRIPTION OF MACHINES

The card punch, or keypunch, machine is the size of a small desk and produces punched cards under keyboard control. It has an alphanumeric keyboard similar to a typewriter keyboard except that the numerals are located in a 3×3 matrix with the top row of the matrix at Keys u, i, and o on the right hand side of the alphabetic keyboard. The shift key serves to shift between alphabetic and numeric information rather than between upper and lower case. The card punch is capable of automatically punching information on one card which it reads from the preceding card. This duplicated information is not liable to keying errors and is not counted in measures of production on a keystroke basis.

The bank proof machine is a device for sorting and totaling bank checks. It has two keyboards: one numeric 3×3 keyboard with a zero bar (similar to an adding machine), and one 4×6 or 4×8 sorting keyboard coded by bank numbers. The operator keys the amount of the check into the numeric keyboard, presses the appropriate sorting key, and drops the check in a slot in the machine. Because of the use of two keyboards and the check handling required, productivity on the bank proof machine cannot be compared directly with card punching.

PRODUCTION RATES

Seven different card punch installations provided sufficient data to make reasonable estimates of the average daily production of the average operator. These estimates range from 56,000 to 83,000 keystrokes per day, with 5 of the 7 in the 66,000 to 74,000 range. The average operator spends about 7 hours per day at the machine. The differences between installations are due largely to differences in the work being punched.

Two card punch installations ran special speed tests of 20–30 minutes on 5 or more operators. Average performance on the 2 tests was 18,000 strokes per hour, or about twice that of the maintained rate over the working day.

Six different banking installations provided sufficient production data to allow estimates of keying rates on bank proof machines. The range in daily production was about 4,350 to 6,600 checks, or about 20,000 to 30,000 keystrokes per day when an average of 4.5 keystrokes is assumed for each check (this does not include the sorting keystroke). The operators spend an average of about 5 hours

per day actually entering checks. The differences in production among installations is largely a matter of sorting requirements and face amounts of the checks handled.

Distributions over Operators

The individual differences among keypunch operators and proof machine operators are similar and may be considered together. Figures 1 and 2 show the distribution of production rates over operators on card punch and bank proof machines, respectively. Both show the number of operators who fall into class intervals of production rates. The effect of excluding operators with less than a year on the job is shown by the dashed line.

Figures 1 and 2 tell the story which seems to hold for large installations. The frequency distribution curves are nearly bell-shaped and symmetrical. Since the distribution curves for production are close to normal it is reasonable to state a standard deviation for their variability. For the experienced keypunch operators of Figure 1 the mean is 87,000 keystrokes per day with a standard deviation of 10,000 keystrokes. For the experienced proof machine operators of Figure 2 the mean production rate is 6,400 checks per day with a standard deviation of 860 checks. Thus, for both keypunch and bank proof operators the standard deviation is approximately 12% of the group mean.

Variability within Operators

In brief, the data of the present study provide the following estimates for within-operator variability in production: week-to-week production averages, for individual operators, fluctuate such that the standard deviation of these averages for each operator is about 6–10% of the group mean production, while monthly averages fluctuate with a standard deviation of about 3–5% of the group mean production (based on variance averaged over operators). This implies that the variability of an operator's production is relatively independent of her mean production level.

This variability within the individual operators contributes to the spread of the distribution in Figures 1 and 2, but it accounts for only a small fraction of the total variance observed between operators. This does not mean that within-operator variability is negligible. Performance measures from keyboard operators do vary sufficiently from day-to-day, week-to-week, and even month-to-month to make consideration of this source of variance important, particularly in evaluating individual performance.

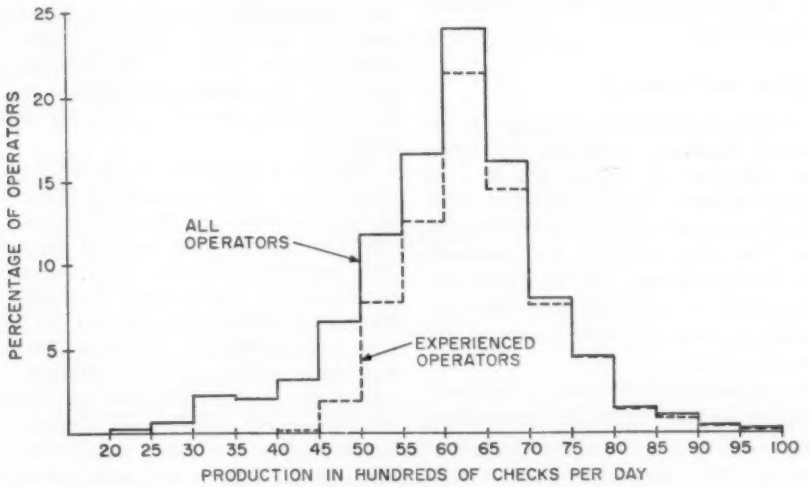

FIG. 2. Distribution of 459 bank proof machine operators according to average daily production rate.

ERROR RATES

The average error estimates, in terms of number of keystrokes per error, range from 1,600 to 4,300 for 4 different card punching installations. In terms of percentage error, this is a range of .02% to .06%. These figures are for errors caught in an independent verifying procedure. No data were available on errors which the operator herself detected and corrected.

Only two of the bank proof machine installations visited keep a record of errors and these records require a few words of explanation. Errors which are not detected by the operator immediately are caught by the operator herself in checking the total of her machine with supplied totals. These are the errors considered here. No errors in the cents position are recorded and some errors in the dimes position are also not counted (usually errors of 10 cents and 20 cents). No sorting errors are counted. Errors which could conceivably be due to poorly written numerals are also not counted. Within these rules, the counted errors average about 3,500 checks per error or .03% error per check. This error is considererably lower than that of any keypunch installation on which we have data. The difference in accuracy may be due to increased emphasis on accuracy, more immediate knowledge of errors, and different criteria for excusing errors in the proof machine installations.

Distribution over Operators

The ranges of individual differences in errors among keypunch operators and proof machine operators are similar and may be considered together. Unlike production data, the units for measuring errors are not straightforward. Two error measures are commonly used by the installations: (*a*) the percentage of items (cards or checks) in errors, and (*b*) the number of items per error (or mean free path between errors).

If the frequency distribution of operators were plotted against intervals of percentage-error, a skewed distribution results with most operators clustered at the low error end. If the number of items per error is used, an even more skewed distribution results with most of the operators clustered at the high error end. Thus, we could apparently show that most operators clustered together in error rate with a few poor operators out in the tail of the distribution, or that most operators clustered together except for a few good operators out in the tail of the distribution. Actually, each of these conclusions is valid for its own measure of error but the resulting skewed distributions do not allow the application of normal curve statistics. In addition, the within-operator variability, as measured by standard deviation of percentage-error, is a linear function of the operator's mean error rate (as described below). Both of these findings suggest a logarithmic transformation. The log transform also has the nice property that the log distribution of errors is the same whether the raw scores are percentage-error or items per error. Since it is difficult to interpret error measures in logarithmic units, it may be better to plot percentage-error on a logarithmic scale. The distributions of undetected errors by operators on card punches and proof machines are plotted in this fashion in Figures 3 and 4. Again, operators with 1 year or more of experience are plotted separately from the entire group.

The difference in performance between the best and worst operators is much more striking in errors than in production. Namely, the ninetieth percentile operator makes 6–10 times as many errors as the tenth percentile operator. The dotted lines indicate the tenth and ninetieth percentile points for each distribution.

There is a similarity in the *shape* of error distributions over operators at different installations which is quite impressive even when the average error rates of the installations are widely different. This similarity can be shown by transforming the error rates for several installations, so that the median error rates are all the same, and superimposing the resulting error distributions. Four sets of data are used to show this similarity: the experienced operators of Figures 3 and 4, the trainees of Figure 3 and a large group of students from a keypunch school after just 1 week of training. The four distributions

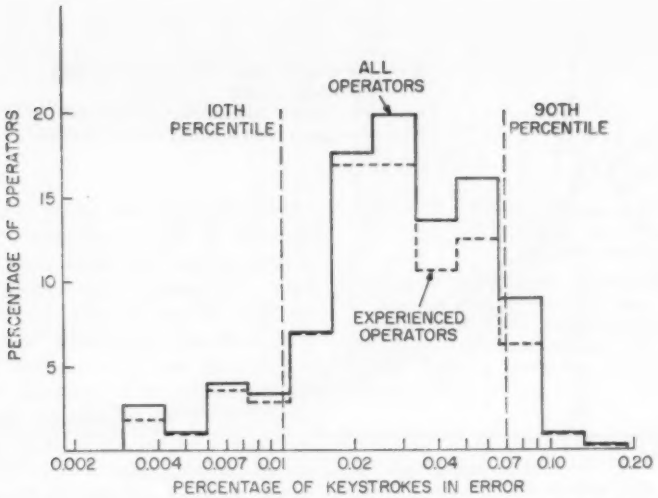

FIG. 3. Distribution of 271 card punch operators according to error rate determined by 10% sample verification over a 1-week period. (Only those errors found in verification are included.)

are shown together after transformation in Figure 5.

The four tasks are all similar—that is, they involve keyboard entry. However, they involve two different keyboards, different types of source documents, and the populations are quite different. So it seems curious that the shape of the distributions is so nearly

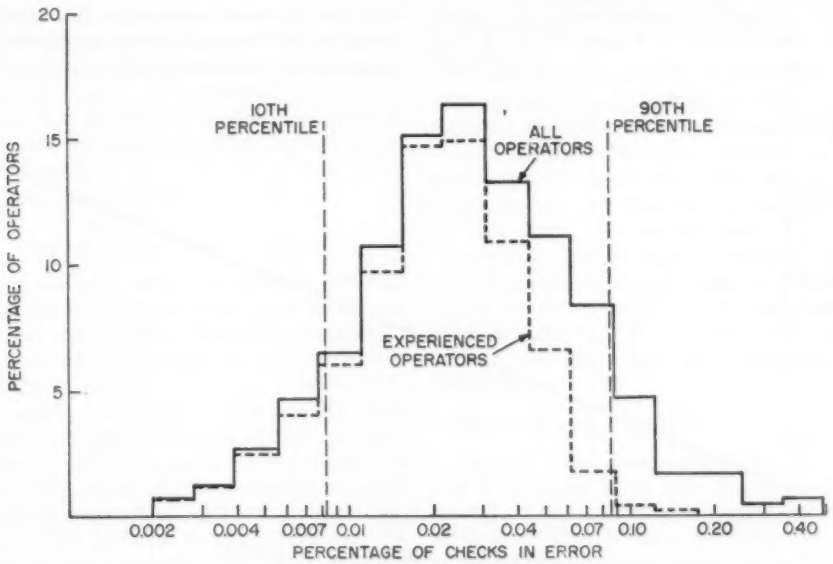

FIG. 4. Distribution of 459 bank proof machine operators according to error rate measured over a 1-month period. (Only those errors found in verification are included.)

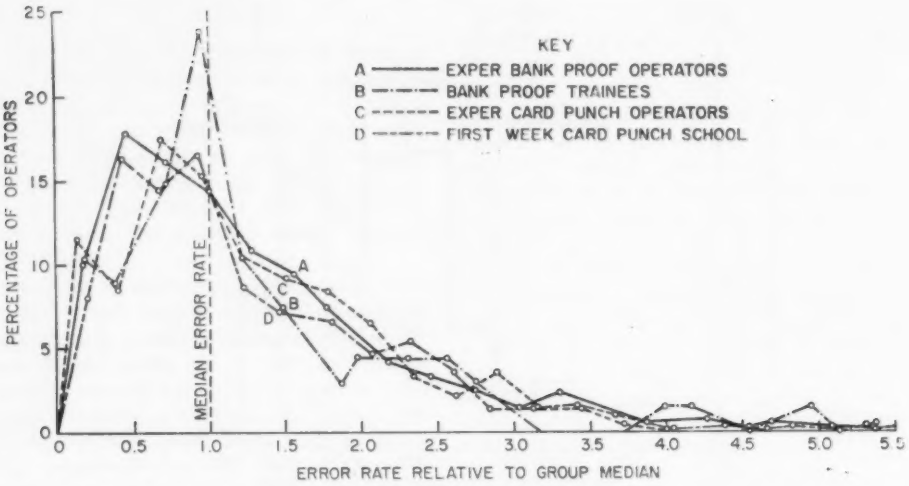
KEY
A EXPER BANK PROOF OPERATORS
B BANK PROOF TRAINEES
C EXPER CARD PUNCH OPERATORS
D FIRST WEEK CARD PUNCH SCHOOL
PERCENTAGE OF OPERATORS
MEDIAN ERROR RATE
ERROR RATE RELATIVE TO GROUP MEDIAN

FIG. 5. Distribution of operators according to error rate for four different groups of operators. (Sampling periods range from 1 hour to 1 month with over 100 operators in each classification.)

the same. Indeed, the only striking difference between the distributions is in average error rates before the data were transformed.

Variability within Operators

As in production distributions, the within-operator variability contributes to the spread of the distributions in Figures 3 and 4, but only accounts for about 10% of the total variance. Unlike the production data, however, the variability of an operator's error rate is not independent of her mean error rate.

The observed within-operator variability as a function of mean error rate is shown in Figure 6 for bank proof operators grouped according to mean error rate. The standard deviation for these operators is obtained by averaging the variance based on scores for 6

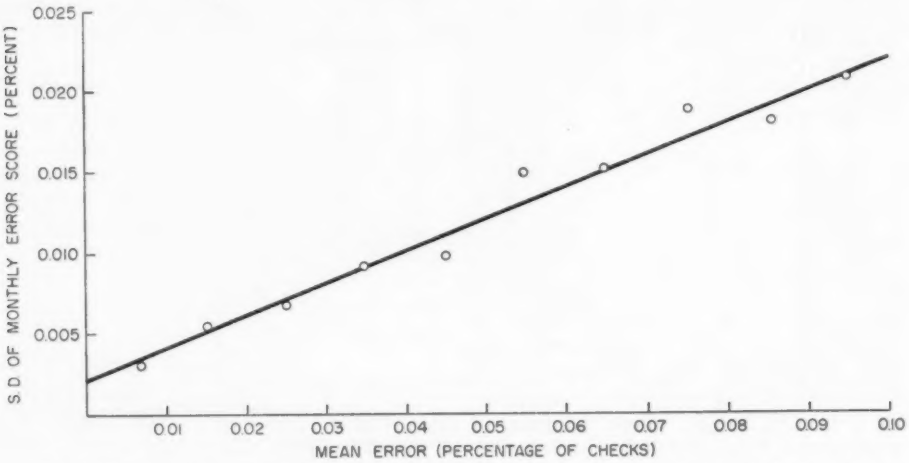

FIG. 6. Within-operator variability of error score as a function of mean error for 459 bank proof machine operators. (Mean error is based on 6 months of data for each operator.)

consecutive months and taking the square root of the average variance. Figure 6 shows that the within-operator standard deviation of percentage-error approximates a linear function of *her* mean error rate.

Poisson Model

Since keying errors are low probability events, it is interesting to ask if observed differences in error rates were due to random fluctuations in observed frequency of a fixed probability event. The answer is that for between-operator differences, the fixed-probability model would predict less than 5% of the observed variance. For within-operator variability the assumption of fixed within-operator error-probability predicts less than half of the observed variance. Thus, the differences in error-probability between operators, and variation of error-probability within operators, are real and cannot be attributed to sampling fluctuations.

RELATION OF SPEED AND ERRORS

Since there is a 2–1 range in speed and a 10–1 range in errors between operators, it is immediately interesting to ask if the fast operators make more errors. Analysis show just the opposite to be true. The operators who are better in speed also make fewer errors, although the correlations are far from perfect. For the keypunch data (one installation) the correlation between speed and errors over all operators is −.42 and for bank proof machines this correlation is −.53.

IMPROVEMENT WITH TRAINING

First we must make a distinction between two methods of obtaining the learning curves. The best way to follow change in performance with practice is to follow the same operators over many months or years of work. Unfortunately, this is not always possible since records are not usually available in this manner. Particularly for data over periods of years, we must depend upon another method. That is, to look at how long each operator has been on the job and to group operators by time intervals according to their starting dates. The intervals must be chosen so as to have a reasonable number of operators in each interval in order to reduce the variability due to individual differences. This consideration dictates a set of intervals increasing in size with increasing months-on-the-job, since normal turnover produces a lesser density of operators at the longer experience end of the curve. The main disadvantage of this second method is that the natural selectivity factors which cause operators to leave must operate differentially to some extent as a function of skill. Learning curves plotted by this second method include the influence of such selectivity factors as well as time on the job.

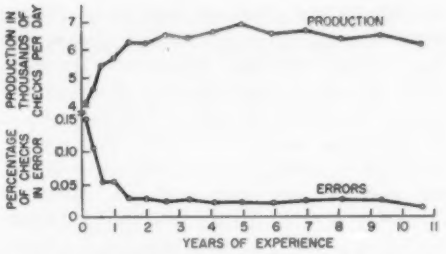

FIG. 7. Improvement in performance as a function of time on the job for 459 bank proof machine operators in one bank. (Plotted points are means for all operators falling within that interval of time on the job. Different operators are represented at each point.)

Production and error rates, as a function of time on the job, are plotted in Figure 7 for one bank proof machine installation using the second method. The mean production and error rates for operators falling within each time interval are plotted against time-on-the-job. This, and similar plots for other installations, clearly shows continued significant improvement in both speed and accuracy during the first year on the job. Improvement after the first year is not clear in all installations and may well be related to job difficulty.

Comparison with Typing

For short-run tests, speed on the card punch averages more than five strokes per second on tasks with no complications. Over the regular working day, averaging over jobs and operators leads to an average production of 2.8 strokes per second during time actually spent on the machine. For 7 hours per day

actually at the machine (out of an 8-hour day) this means about 70,000 strokes per day. Translated into typewriter language, these figures are equivalent to 60 words per minute for short runs with a daily average of 34 words per minute for 7 hours of typing. It should be noted, however, that these figures are averages over time and that operators often key easy on highly practiced combinations (e.g., "the") at rates pressing the limiting speed of the machine (18–20 strokes per second).

REFERENCES

COLVER, G. (Ed.) *A guide to office clerical time standards*. Detroit, Mich.: Systems and Procedures Association, 1960.

HILLIX, W. A., & COBURN, R. Human factors in key set design. *USN Electron. Lab. res. Rep.*, 1961, No. 1023.

RAHE, H. *Typewriting research index: 1900–1954*. Carbondale, Ill.: 807 Twisdale Avenue, 1954.

WEST, L. J. Review of research in typewriting learning with recommendations for training. *USAF Train. Res. Cent. tech. Note*, 1957, No. 57-69.

(Received December 12, 1961)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 409–416

THE VALIDITY OF TWO-SCALE PATTERN INTERPRETATION ON THE CALIFORNIA PSYCHOLOGICAL INVENTORY

ALFRED B. HEILBRUN, JR., JOHN L. DANIEL, LEONARD D. GOODSTEIN, RICHARD R. STEPHENSON, AND JOHN O. CRITES

University of Iowa

The validity of 2-scale pattern interpretation for the CPI was assessed within a college counseling service population. Groups representing high and low score combinations on the Dominance and Good Impression scales were constituted, and the self-descriptive adjectives from the Adjective Check List which differentiated these groups from a representative counseling service group were determined. These clusters of self-descriptive terms were generally consistent with the personality types which would be anticipated, based on the combined meanings of high and low scores on the individual scales as evidenced by high matching accuracy of expert judges. These results provide gross validational support for 2-scale pattern interpretation on the CPI.

The availability of a valid multidimensional personality test for use in the assessment of the relatively normal (nonpsychiatric) population of college students who seek counseling is doubly important. Not only can the personality inferences be of substantial benefit in formulating early hypotheses about the client, but urgently-needed counseling research involving personality variables can be implemented. A series of investigations is currently being conducted to assess the validity of one relatively new personality test, the California Psychological Inventory (CPI) (Gough, 1957), with college counseling service subjects. The initial interest in this instrument was based on two of its characteristics. For one, its 18 scales measure a wide range of interpersonal behaviors relevant to the relatively normal clientele of a counseling service. For the second, many of its scales were developed empirically, allowing a somewhat higher degree of confidence in their ultimate validation.

Gough (1957, pp. 14–16) recommends and describes a five-step procedure for CPI profile interpretation, and the validational research of the current investigators has followed the same order. The first step in interpretation is an appraisal of overall elevation. Since the scales are nearly all keyed so high scores indicate positive behaviors and low scores suggest negative attributes, the overall profile elevation should be positively related to level of adjustment. Goodstein, Crites, Heilbrun, and Rempel (1961) found support for this assumption by showing the expected profile elevation differences between three groups of college students clearly differing in level of adjustment (i.e., counseling service personal adjustment clients, counseling service vocational-educational clients, and college normals).

Gough has grouped his 18 scales into four classes, and he suggests that the test interpreter should, as a second step, note the differential elevations of these scale clusters. His groupings include: (*a*) "measures of poise, ascendancy, and self-assurance" (Dominance, Capacity for Status, Sociability, Social Presence, Self-Acceptance, and Sense of Well-Being); (*b*) "measures of socialization, maturity, and responsibility" (Responsibility, Socialization, Self-Control, Tolerance, Good Impression, and Communality); and (*c*) "measures of achievement potential and intellectual efficiency" (Achievement via Conformance, Achievement via Independence, Intellectual Efficiency); and (*d*) "measures of intellectual and interest modes" (Psychological-Mindedness, Flexibility, Femininity). Although not strictly a validational study, Crites, Bechtoldt, Goodstein, and Heilbrun (1961) did test the adequacy of Gough's groupings by a factor analysis of the 18 scales, the subjects again being both client and nonclient college students. In the main, Gough's Groups *a* and *b* were extracted intact from the matrix of intercorrelations, but there was considerably less support for Gough's Groups *c* and *d*.

The third step in Gough's recommended

approach to CPI interpretation is to note the high and low scale scores as one moves toward greater specificity in his inferences. Included in this mode of interpretation is the technique of two-scale pattern analysis[1] which is presented as a "tentative" method in the manual. Two-scale pattern analysis involves the classification of individuals into four groups as a function of their concomitant deviation from some baseline on two selected scales. Thus, some would be high on both Scales X and Y, some high on X but low on Y, some low on X and high on Y, and some low on both X and Y. To be of any value, two-scale pattern analysis must provide the interpreter with valid inferences which are not available from the scores on either X or Y when considered independently. If it could be shown that the CPI could generate such valid and unique statements based upon relative elevations of scale pairs, the breadth of the test's usefulness would be considerably increased. The present study is concerned with assessing the validity of two-scale pattern analysis with the CPI.

METHOD

California Psychological Inventory Scales

The CPI is comprised of 480 true-false items whose content primarily samples behaviors of a social interactional nature but also includes attitudes toward achievement and authority. The two scales of the 18 included in the CPI which were selected for the pattern analysis were the Dominance (Do) and Good Impression (Gi) scales. These particular scales were chosen because they were previously found to represent best the two major factor analytically derived clusters (Crites et al., 1961), having high loadings on their respective factors, and showing the least item overlap with scales in other factors. Gough (1957, p. 12) considers Do as a measure of "factors of leadership ability, dominance, persistence, and social initiative" and provides the following descriptions of the high and low Do scorer:

1. High Do scorer—"aggressive, confident, persistent, and planful; as being persuasive and verbally fluent; as self-reliant and independent; as having leadership potential and initiative."
2. Low Do scorer—"retiring, inhibited, commonplace, indifferent, silent, and unassuming; as being slow in thought and action; as avoiding situations of tension and decision; and as lacking in self-confidence."

The Gi scale is represented by Gough (1957, p. 12) as a measure "to identify persons capable of creating a favorable impression, and who are concerned about how others react to them." The descriptions of high and low Gi scorers are:

1. High Gi scorers—"cooperative, enterprising, outgoing, sociable, warm, and helpful; as being concerned with making a good impression; and as being diligent and persistent."
2. Low Gi scorer—"inhibited, cautious, shrewd, wary, aloof, and resentful; as being cool and distant in their relationships with others; and as being self-centered and too little concerned with the needs and wants of others."

Considering the implications of high and low scores on both scales, it would be anticipated that certain unique behavioral attributes would emerge when taken in combination. The anticipated pattern typologies are:

1. High Do-high Gi—This person will assert himself in interpersonal relationships so as to assume supraordinate roles but will do so tactfully and in such a way as to initiate and sustain the high personal regard of those who are subordinated to him. His satisfactions stem from initiating actions which do things *for* others ("benevolent leader" type).
2. High Do-low Gi—This person will assert himself in interpersonal relationships so as to assume supraordinate roles but will do so aggressively and without concern for the personal feelings of those whom he subordinates. His satisfactions stem from initiating actions which do things *to* others ("malevolent dictator" type).
3. Low Do-high Gi—Conflicted about social participation, this person will seek social relationships which do not require initiative or responsibility but which will put him in a position to obtain social rewards. This is the person who blends into his social background. ("He always seems like such a nice person—wonder what his name is?" type).
4. Low Do-low Gi—This person is socially aloof, even withdrawn and resentful, making no pretenses about being interpersonally active. Self-centeredness brings few rewards except to escape the threat of social demands for which he lacks motivation or skills ("disgruntled recluse" type).

Subjects

The basic pool of subjects for this study was 947 clients at the University Counseling Service, University of Iowa. This number included 638 males and 309 females and was divided in about a 5:1 ratio of vocational-educational to personal adjustment counseling cases. Each had been given the CPI as part of the standard intake battery.

[1] A similar interpretative technique termed "configural scoring" was earlier suggested by Meehl (1950) at the item level on objective personality questionnaires.

Procedure

Assignment to the four pattern classes was instrumented by first estimating male and female college means and standard deviations for the Do and Gi scales. Gough (1957) had provided college norms which indicate male means to be 53 and 47, respectively, and female means to be 53 and 48, respectively. The sigma is 10 in each case. High or low in the present study was defined as deviating more than a half sigma (> 5) above or below the college mean. Based on this definition, the numbers out of 638 total in the four male groups were: high Do-high Gi, $N = 31$; high Do-low Gi, $N = 35$; low Do-high Gi, $N = 27$; low Do-low Gi, $N = 129$. Of the 309 total females, the high Do-high Gi group included 22; high Do-low Gi, 18; low Do-high Gi, 15; and low Do-low Gi, 64.

Stratified samples of "counseling clients in general" relative to concomitant elevations of the Do and Gi scales were then constituted as reference (comparison) groups. The male reference group included 110 clients whose Do and Gi scores did not place them into any of the four pattern groups and, in addition, proportions of subjects from the four pattern groups based upon the prevalence of these two-scale patterns in male counseling clients. For example, 129 out of 638 (20%) male clients

TABLE 1

SELF-DESCRIPTIVE TERMS DIFFERENTIATING HIGH DOMINANT-HIGH GOOD IMPRESSION MALE AND FEMALE CLIENTS FROM REPRESENTATIVE CLIENT GROUP

Male high Do-high Gi More:	Male high Do-high Gi Less:	Female high Do-high Gi More:	Female high Do-high Gi Less:
Active	Anxious	Contented	Anxious
Adaptable	Careless		Civilized
Adventurous	Complaining		Confused
Ambitious	Confused		Conservative
Confident	Cynical		Curious
Determined	Defensive		Defensive
Foresighted	Dissatisfied		Dependent
Generous	Distractable		Dissatisfied
Individualistic	Dreamy		Easy-going
Industrious	Emotional		Emotional
Inventive	Evasive		Fair-minded
Jolly	Excitable		Forgiving
Original	Hasty		Fussy
Patient	High strung		Good-natured
Progressive	Hurried		Healthy
Resourceful	Immature		Lazy
Self-confident	Impatient		Leisurely
Spunky	Irritable		Loyal
Strong	Moody		Moody
Talkative	Nervous		Pessimistic
Wise	Pessimistic		Precise
	Prejudiced		Restless
	Self-pitying		Serious
	Silent		Silent
	Spendthrift		Simple
	Suspicious		Sincere
	Temperamental		Slow
	Tense		Soft-hearted
	Touchy		Stubborn
	Unambitious		Submissive
	Unstable		Talkative
	Withdrawn		Timid
	Worrying		Tolerant
			Unassuming
			Withdrawn

TABLE 2

SELF-DESCRIPTIVE TERMS DIFFERENTIATING HIGH DOMINANT–LOW GOOD IMPRESSION MALE AND FEMALE CLIENTS FROM REPRESENTATIVE CLIENT GROUPS

Male high Do-low Gi More:	Male high Do-low Gi Less:	Female high Do-low Gi More:	Female high Do-low Gi Less:
Adventurous	Commonplace	Assertive	Adaptable
Aggressive	Conservative	Commonplace	Calm
Aloof	Contented	Complicated	Conscientious
Argumentative	Dull	Cynical	Cooperative
Assertive	Easy-going	Daring	Considerate
Clever	Modest	Demanding	Dependable
Confident	Peaceable	Disorderly	Honest
Demanding	Shy	Dominant	Independent
Egotistical	Silent	Egotistical	Quiet
Enterprising	Simple	Enterprising	Sincere
Forceful	Quiet	Fault-finding	Slow
Humorous		Fearful	Timid
Impulsive		Forceful	Undependable
Intolerant		Friendly	Withdrawn
Mannerly		Immature	
Outspoken		Ingenious	
Polished		Mischievous	
Rational		Opinionated	
Resourceful		Opportunistic	
Self-confident		Optimistic	
Sentimental		Original	
Severe		Outspoken	
Sharp-witted		Quick	
Show-off		Sarcastic	
Shrewd		Shrewd	
Spontaneous		Sociable	
Spunky		Sophisticated	
Stubborn		Spontaneous	
Suspicious		Spunky	
Talkative		Wary	
Versatile			
Witty			

showed the low Do-low Gi pattern; accordingly, 21 (20% of 110) low Do-low Gi subjects were included in the male reference group. The stratified male reference group and a similarly constituted stratified female reference group included 138 and 150 subjects, respectively.

In addition to the CPI, each client had also been administered the Gough Adjective Check List (ACL) (Gough & Heilbrun, in press) as part of the counseling service intake battery, which test requires the person to check those adjectives deemed self-characteristic. The validational procedure employed in the present study was to determine which of the 300 adjectives were differentially endorsed as self-characteristic by each of the pattern groups relative to the appropriate reference group; then, the psychological meanings of these adjective clusters were related to the pattern typologies implied by the combined meanings of the individual Do and Gi scales.

More specifically, 35 judges (including 15 PhD psychologists from southern Iowa and 20 advanced graduate students in psychology at the University of Iowa) were asked to match the adjective clusters to the typologies. Male and female analyses were conducted separately since it remained an empirical question whether the traits measured by the Do and Gi scales or their combined patterns would be manifested interpersonally in the same ways for both sexes. Thus, each judge matched the four adjective clusters distinguishing the pattern groups with the four pattern typologies independently for males and females. The percentage of accurate matchings represents the validating statistic.

RESULTS AND DISCUSSION

Chi square analysis was used to determine which ACL adjectives were differentially endorsed by each pattern group relative to a reference group with the $p < .05$ level required for significance. The self-descriptive terms which differentiated the high Do-high Gi male and female groups from their reference groups are given in Table 1. The predicted typology is well supported in the case of high Do-high Gi males, since outgoing leadership qualities as well as a regard and concern for others are consistent attributes of this class of clients. The high Do-high Gi females do not present as clear-cut a pattern of endorsement, however. There is a suggestion of interpersonal dominance in their denial of passivity, but there is also an unexpected denial of attributes which serve to elicit favorable social responses.

The adjectives distinguishing high Do-low Gi clients from the reference groups are presented in Table 2. The descriptive terms for the high Do-low Gi males are quite compatible with the predicted pattern typology. The outgoingness, initiative, and leadership qualities are intermixed with antagonistic terms like aggressive, argumentative, assertive, egotistical, forceful, and outspoken, none of which characterized the high Do-high Gi males. The high Do-low Gi female clients also provide a pattern of self-descriptions strikingly consistent with typology both in terms of the anticipated dominating qualities and their aggressiveness toward those whom they would presumably dominate.

Table 3 gives the adjectives differentiating the low Do-high Gi male and female clients from their respective reference groups. Although there were considerably fewer adjectives distinguishing this type of client, the predicted typology of the nonforceful, undistinguished person who avoids interpersonal responsibilities but seeks social approval emerges for both sexes. It is interesting that the constriction implied by this typology appears in the guise of a general cautiousness for the males (painstaking, stern, thrifty, less hasty) and as a lack of emotional involvement for females (unemotional, less affectionate, less cheerful, less serious, less tense).

The distinguishing personological terms for the low Do-low Gi clients are found in Table 4. Inspection of these lists provides a convincing array of terms portraying the withdrawn, socially inept, resentful male or female and thus provides support for the pattern typology.

The percentage of the 35 judges accurately matching each adjective cluster with the ap-

TABLE 3

SELF-DESCRIPTIVE TERMS DIFFERENTIATING LOW DOMINANT–HIGH GOOD IMPRESSION MALE AND FEMALE CLIENTS FROM REPRESENTATIVE CLIENT GROUPS

Male low Do-high Gi More:	Male low Do-high Gi Less:	Female low Do-high Gi More:	Female low Do-high Gi Less:
Forgetful	Boastful	Patient	Affectionate
Friendly	Confused	Quiet	Aggressive
Painstaking	Dependent	Unemotional	Cheerful
Stern	Egotistical		Clever
Strong	Enthusiastic		Idealistic
Thrifty	Flirtatious		Lazy
	Forceful		Mannerly
	Hasty		Outgoing
	Methodical		Outspoken
	Show-off		Serious
	Spendthrift		Suggestible
	Sutbborn		Tense
	Suspicious		
	Touchy		

TABLE 4

SELF-DESCRIPTIVE TERMS DIFFERENTIATING LOW DOMINANT–LOW GOOD IMPRESSION MALE AND FEMALE CLIENTS FROM REPRESENTATIVE CLIENT GROUPS

Male low Do-low Gi More:	Male low Do-low Gi Less:	Female low Do-low Gi More:	Female low Do-low Gi Less:
Absent-minded	Active	Absent-minded	Aggressive
Awkward	Adaptable	Aloof	Alert
Bitter	Affectionate	Anxious	Capable
Careless	Aggressive	Bitter	Calm
Changeable	Alert	Careless	Cheerful
Complaining	Ambitious	Changeable	Determined
Complicated	Assertive	Complaining	Efficient
Confused	Calm	Confused	Energetic
Despondent	Capable	Defensive	Enthusiastic
Dissatisfied	Cheerful	Despondent	Good-natured
Distrustful	Clear-thinking	Dissatisfied	Industrious
Dreamy	Clever	Dreamy	Initiative
Emotional	Confident	Egotistical	Inventive
Gloomy	Contented	Fearful	Natural
High-strung	Cooperative	Gloomy	Out-going
Immature	Courageous	Healthy	Patient
Impatient	Determined	Immature	Persevering
Inhibited	Dominant	Impatient	Pleasant
Interests narrow	Efficient	Indifferent	Poised
Intolerant	Energetic	Infantile	Progressive
Irritable	Enterprising	Inhibited	Quiet
Lazy	Enthusiastic	Intolerant	Realistic
Moody	Forceful	Irresponsible	Reliable
Nervous	Foresighted	Irritable	Resourceful
Preoccupied	Friendly	Lazy	Robust
Quitting	Healthy	Meek	Self-confident
Reckless	Helpful	Moody	Sociable
Restless	Honest	Nervous	Stable
Retiring	Humorous	Pessimistic	Talkative
Shy	Industrious	Quitting	Zany
Silent	Ingenious	Rebellious	
Slow	Initiative	Resentful	
Tactless	Insightful	Retiring	
Temperamental	Interests wide	Selfish	
Timid	Logical	Self-punishing	
Touchy	Masculine	Sexy	
Unintelligent	Mature	Sulky	
Wary	Methodical	Tactless	
Weak	Natural	Temperamental	
Withdrawn	Opportunistic	Tense	
Worrying	Optimistic	Touchy	
	Persistent	Unambitious	
	Pleasant	Undependable	
	Poised	Withdrawn	
	Practical		
	Precise		
	Progressive		
	Realistic		

Table 4—Continued

Male low Do-low Gi More:	Male low Do-low Gi Less:	Female low Do-low Gi More:	Female low Do-low Gi Less:
	Reasonable		
	Relaxed		
	Reliable		
	Resourceful		
	Responsible		
	Self-confident		
	Self-controlled		
	Shrewd		
	Sincere		
	Sophisticated		
	Stable		
	Steady		
	Stern		
	Tactful		
	Thorough		
	Thoughtful		
	Thrifty		
	Trusting		
	Understanding		
	Uninhibited		
	Unscrupulous		
	Unselfish		
	Versatile		
	Wholesome		
	Witty		

propriate typology is given in Table 5. The typologies anticipated on the basis of the combined meanings of high and low scores on the Do and Gi scales were systematically and accurately matched with the patterns of independently obtained self-descriptions for both males and females as evidenced by high degrees of accuracy ranging from 74% to 97% for the eight matchings (average percentage of accuracy = 88). These figures were sufficiently high so as to obviate tests for statistical significance, since random matching of adjective clusters to typologies should result in but 25% accuracy. Although the procedure used in this study for validating two-scale pattern CPI interpretation is crude, amounting as it does to matching experts' judgments of global meanings for empirically-derived clusters of self-descriptive terms with anticipated two-scale pattern typologies, there would probably be little argument with the conclusion that the results are generally positive.

The investigators further recognize that the findings of the present study represent systematic relationships between two sets of self-descriptive behaviors. Although it might be argued that such findings should be constructed as bearing upon consistency of self-description (reliability) rather than validity of two-scale pattern analysis, it is felt that either interpretation of the results should enhance confidence in CPI pattern analysis.

TABLE 5

PERCENTAGE OF JUDGES CORRECTLY MATCHING ADJECTIVE CLUSTERS WITH APPROPRIATE INTERACTIONAL TYPOLOGIES

(N = 35)

	Interactional typology			
Sex judged	High Do-high Gi (%)	High Do-low Gi (%)	Low Do-high Gi (%)	Low Do-low Gi (%)
Male	91	94	89	94
Female	74	91	74	97

Since one of the scales (Gi) chosen for investigation was derived as a validity check for the CPI to detect persons trying to portray themselves in an unduly favorable light, the question may be raised as to whether the findings merely represent a distinguishing of high and low dominant groups who differ in the degree to which they are willing to portray themselves unfavorably on the test. Such an interpretation presumes that the tendency to portray oneself favorably is not a generalized trait but rather is specific to a psychometric situation. Gough's (1957) interpretation of elevations on Gi is that "scores somewhat above average . . . are indications of favorable attitudes and conscientious efforts to fit in and to adapt. Very high scores raise the possibility of the test 'faking,' or of an undue concern with making a good impression" (p. 19). In either case, Gough's remarks seem best interpreted as suggesting a generalized trait of seeking social approval for one's behavior, and there is now evidence supporting this position. Two studies have found that giving socially desirable responses in a specific testing situation is not a source of error in predicting nontest behaviors (Heilbrun, 1962; Heilbrun & Goodstein, 1961b). Heilbrun and Goodstein (1961a) report that the social desirability values attached to test trait statements are themselves predictive of nontest behavior. Finally, it was found that the degree to which a person's test responses are related to social desirability is predictive of the degree to which his social behavior is oriented toward creating a favorable impression (Marlow & Crowne, 1961). Thus, there is good reason to presume that the trait measured by scales such as Gi is a generalized one and that the Do-Gi typologies have predictive significance beyond the immediate testing situation as well.

REFERENCES

CRITES, J. O., BECHTOLDT, H. P., GOODSTEIN, L. D., & HEILBRUN, A. A factor analysis of the California Psychological Inventory. *J. appl. Psychol.*, 1961, **45**, 408–414.

GOODSTEIN, L. D., CRITES, J. O., HEILBRUN, A. B., & REMPEL, P. P. The use of the California Psychological Inventory in a university counseling service. *J. counsel. Psychol.*, 1961, **8**, 147–153.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.

GOUGH, H. G., & HEILBRUN, A. B. *Joint manual for the Adjective Check List and the Need scales for the ACL*. Palo Alto, Calif.: Consulting Psychologists Press, in press.

HEILBRUN, A. B. Social desirability and the relative validities of two achievement scales. *J. consult. Psychol.*, 1962, **26**, 383–385.

HEILBRUN, A. B., & GOODSTEIN, L. D. Consistency between social desirability ratings and item endorsement as a function of psychopathology. *Psychol. Rep.*, 1961, **8**, 69–70. (a)

HEILBRUN, A. B., & GOODSTEIN, L. D. Social desirability: Error or predictor variable? *J. Psychol.*, 1961, **51**, 321–329. (b)

MARLOW, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, **25**, 109–115.

MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, **14**, 165–171.

(Received January 18, 1962)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 417–419

RE-EVALUATION OF A WEIGHTED APPLICATION BLANK FOR OFFICE PERSONNEL

PAUL F. WERNIMONT

University of Minnesota

The validity of weighted application blank scores in predicting length of service of female office personnel of a large midwestern manufacturing company showed a slow but consistent decline from 1954 to 1959. After reweighting the application blank variables by the usual procedure, a Pearson r of .39 was obtained between WAB scores and tenure on a cross-validation sample of 208 individuals. Only 3 variables (high proficiency at shorthand; did not leave last job because of pregnancy, marriage, sickness, or home problems; and will begin work on the new job 1 week or more from now) retained their predictive validity from 1954 to 1959. Some of the variables were assigned different weights, and others were given weights for the 1st time. It was recommended that weights assigned to application blank responses be reviewed at least every 3–5 years.

In recent years, the weighted application blank has come into increasing prominence as a selection device in industry. This trend reflects, in part, the desire to express selection methods in quantitative terms, and in part, the increasing recognition that the application blank has potential value beyond its use as anecdotal information. England (1961) presents a detailed account of the methodology of the development of a weighted application blank, as well as an extensive list of references in this area. Of those studies using turnover as the criterion, Pearson *r*'s near .40 have generally been reported.

Reports of long-term studies of the validities of weighted application blanks are virtually nonexistent in the literature. Perhaps the need for subsequent review of such validities is too obvious to require mentioning. At any rate, when turnover is used as the criterion, it is certainly not to be expected that the same application blank information, for instance, highest past salary, should indefinitely retain its predictive validity for new job applicants.

The purpose of this report is to present the results of a re-evaluation of the weighted application blank developed by Kirchner and Dunnette (1957) on the basis of female office personnel hired in 1954. The female office personnel included in this study held essentially the same types of jobs as those included in the earlier study, and in both instances, were employees of the same large midwestern manufacturing company. Dunnette, Kirchner, Erickson, and Banas (1960) have already reported 1- and 2-year follow-ups on this same set of application blank weights. Their results indicated that the predictive validity, while fairly high, had decreased somewhat from that of the initial study.

Method

As a starting point for this study, a sample of approximately one-fourth of the weighted application blank scores of the female office employees hired in 1959, for the home office, was correlated with tenure. A tetrachoric *r* of .07 was found, indicating that the predictive value of the variables weighted in 1954 had deteriorated considerably. It was decided to attempt a reweighting of the application blank responses.

The application blanks for all new female office workers hired during 1959 were divided into two groups and systematically reviewed. Those hired during the first half of the year were used for the main part of the study, while those hired during the last half of 1959 comprised the holdout, or cross-validation group. As in the initial study, length of service was chosen as the criterion. The primary sample was grouped as follows: 5 months and under ($N = 75$), 6–18 months ($N = 53$), 19 months and over ($N = 83$). Tenure was divided in this way because of clear breaks in the distribution at the 6- and 19-month points.

Using the traditional approach, application blanks of the long-term group were compared with those of the short-term group on a total of 37 variables. Of these, 16 were found to differentiate, on the basis of at least a 10% difference in the frequency of responses, between the two groups. Weights of −1, 0,

and 1 were then assigned to responses for each variable. Total scores were not computed for the application blanks used in deriving the weights, but all three (long-, medium-, and short-term) cross-validation groups were scored on the 16 weighted variables, and total scores were obtained. This cross-validation sample was grouped as follows: 5 months and under ($N = 71$), 6–18 months ($N = 61$), 19 months or more ($N = 76$).

RESULTS AND DISCUSSION

From Table 1, it may be observed that the new total scores do effectively differentiate between the long- and short-term groups, with the long-term group receiving, on the average, higher scores than did the short-term people. The difference between the means is highly significant statistically, in this, the cross-validation sample. The relatively small overlap found provides impressive evidence that these 16 weighted variables are effective in differentiating the long- from the short-term employees. It was found that 84% of the long-term group equal or exceed the median of the short-term group, and none of the long-term group scored below a -2, or Q_1 of the short-term group. A product-moment r of .39 was obtained when length of service was correlated with these new weighted application blank total scores. This coefficient compares favorably with those found in other studies of this kind. Since some of the variables used in deriving the old application blank scores were also included in the new set of weighted variables, it is evident that the office employees had already been partially screened. This situation would, of course, tend to produce an apparent shrinkage of the obtained validity coefficient.

From the results of this study, it is readily apparent that scoring weights based on application blank responses need to be re-evaluated from time to time. There were some variables in this study that had to be assigned different weights from those assigned in the earlier study, and there were several variables that operated in 1959 to differentiate long-term from short-term employees that did not do so in 1954. There were also a few variables that still held up, and in the same way, from the 1954–59 groups.

The variables which were weighted differently include length of time in last position, number of past positions, age, and other office

TABLE 1

QUARTILES AND MEANS OF WEIGHTED SCORES FOR LONG- AND SHORT-TERM OFFICE EMPLOYEES HIRED IN LAST HALF OF 1959
(Cross-validation group)

Group	Q_1	Median	Q_3	M	SD	N
Short	−2	1	3	.80	3.72	71
Long	1	4	6	3.84	3.44	76

training. The new weights for these variables indicate that the long-term employee in 1959 was an older, more experienced individual than her 1954 counterpart. Among the new variables which were found to differentiate on the tenure criterion were average high school grade, type of work in last position, length of time expected to stay with the company, salary expected, position desired, and an evidence of leadership statement. It appears that the long-term employee, having been dissatisfied with her salary and chances to advance in her previous job, took the advice of relatives to join this company, where she expects to advance in both salary and position. There were three variables which consistently held up in the same way from 1954 to 1959, including the 1956 follow-up study. These variables are high proficiency at shorthand; did not leave last job because of pregnancy, marriage, sickness, or home problems; and will begin work on the new job 1 week or more from now. This group of variables suggests that the long-term female job applicant is still working on her previous job, but intends to leave it, if the new job meets her expectations for advancement.

Tetrachoric r's between total application blank scores and tenure have been computed by Dunnette et al. (1960) as follows: 1954—.74, 1955—.61, 1956—.38. The 1955 group was a cross-validation group, and the 1956 correlation was based on a follow-up study. The tetrachoric r of .07 obtained in the present study, based on the same weights, indicates that there was a slow but continuous deterioration of the ability of the weighted application blank scores to differentiate long- from short-term employees. After a period of 5 years, the results of using these scores are little better than chance. Scores based on the revised .weights and set of variables, when

computed on the cross-validation group, show a tetrochoric *r* of .57 with tenure.

The maximum time that a given set of weights can safely be used without review probably depends upon a large number of factors, among them economic conditions, locale, type of industry, selection ratio, the criterion used, etc. It seems safe to say, however, that weights assigned to application blank responses ought to be reviewed at least every 3–5 years.

REFERENCES

DUNNETTE, M. D., KIRCHNER, W. K., ERICKSON, J. R., & BANAS, P. A. Predicting turnover of female office employees. *Personnel Admin.*, 1960, **23**, 45–50.

ENGLAND, G. W. *Development and use of weighted application blanks.* Dubuque, Ia.: Wm. C. Brown, 1961.

KIRCHNER, W. K., & DUNNETTE, M. D. Applying the weighted application blank technique to a variety of office jobs. *J. appl. Psychol.*, 1957, **41**, 206–208.

(Received January 25, 1962)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 420–424

MOTION PREDICTION AS A FUNCTION OF TARGET SPEED AND DURATION OF PRESENTATION [1]

EARL L. WIENER

University of Miami

This study investigated the ability of Ss to predict the future position of a moving target after the target disappeared. Target speed, duration of target exposure, and S's mode of responding to the visible target were varied. The performance measure was the absolute deviation from the correct target position at the end of 9 sec., converted to error relative to target speed. Results show: (a) no significant differences resulting from mode of response (tracking vs. monitoring), order of presentation, duration of presentation, or speed-duration interaction; (b) significant learning effect from session to session ($p < .01$); and (c) an increase in relative error, in an inverse relation to target speed ($p < .01$). It is concluded that a human operator may be able to make motion predictions equally as well with minimal as with maximal exposure to target input; only target speed exerts an influence on prediction accuracy.

Prediction of the future position of a moving target is an essential task of the human operator in any system in which he monitors or controls a dynamic, continually-varying process. Prediction may be defined as an extrapolation to a future position from current information on the state of the system and probable changes in that state. The efficiency of a system is, to a large degree, dependent on the success with which an operator can anticipate its future state.

Prediction is especially important when input information or feedback loops are subject to degradation as, for instance, in the occurrence of functional breakdown of equipment. In such an emergency, if the operator is to continue functioning in the system, he must make prediction based on prebreakdown inputs. This is particularly true of radar displays, where even with the equipment functioning properly, the target under surveillance may temporarily disappear in scope clutter.

Several authors (Bowen & Woodhead, 1953, 1955; McGuire, 1956; Manglesdorf, 1955; Manglesdorf & Fitts, 1954a, 1954b; Schipper & Versace, 1956) have investigated man's ability to make predictions of future position on a static display. Generally these authors have reported that accuracy of prediction varies inversely with the distance to be extrapolated, but is unaffected by the length of the initiating target which represents speed.[2] Using a dynamic display, Gottsdanker (1952a, 1952b, 1955) demonstrated a phenomenon which he called "rate smoothing": the subjects (*Ss*) tended to underestimate accelerating targets and overestimate decelerating targets. Accuracy was very high on constant rate targets.

The present study investigated the effect of target speed, duration of exposure, and mode of response on prediction of future position of constant-rate targets following disappearance of the target.

METHOD

Apparatus

The apparatus was an adaptation of that used by Gottsdanker (1952a). A moving target was produced by driving a chart with a diagonal pencil line under a 2 mm. slit in an aluminum plate. The plate was

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(616)-3612 with the Ohio State University Research Foundation, and was conducted in the Laboratory of Aviation Psychology. This report is based upon the thesis submitted in partial fulfillment of the requirements of an MA degree at Ohio State. The author is indebted to G. E. Briggs who served as his adviser.

[2] It is interesting that one of the first pilot selection devices involved motion prediction with a static display. Stratton, McComas, Coover, and Bagby (1920) report the use of such a device in selecting World War I aviators. The subject was required to estimate where one branch of a parabolic curve would intersect a horizontal line. This was thought to be an analog of the landing task. The correlation between scores of this test and flight instructors' ratings was .05.

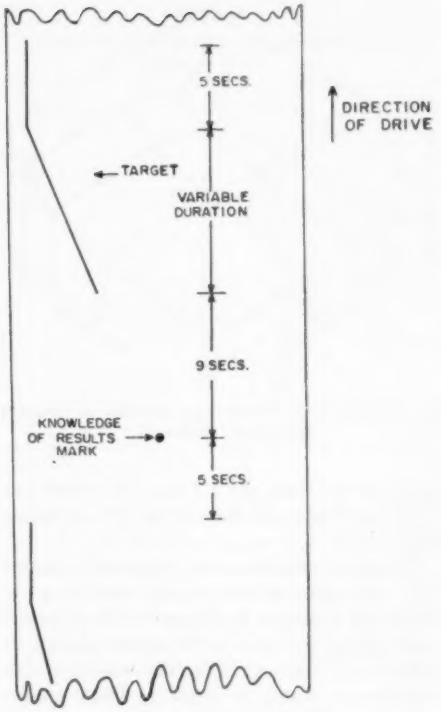

FIG. 1. A sample stimulus.

flush-mounted in the top of a plywood box which held the drive mechanism, an Esterline-Angus Model 38-D chart drive. The targets were drawn on blank Esterline-Angus chart paper. The target appeared as a dot moving from left to right through the slit.

Stimuli

Four durations of exposure (2, 4, 8, and 16 seconds) and four speeds (.5, 1, 2 and 4mm/sec) were combined forming 16 unique combinations of stimulus lines. The various speeds were produced by varying the slope of the target line and the durations by the length of the line.

Each target appeared in the extreme left of the slit remaining stationary for 5 seconds before beginning its left-to-right travel. Nine seconds after the disappearance of the target a red dot appeared on the extension of the target line. This indicated the end of a trial and provided knowledge of results at the terminal point. Five seconds later a new target appeared at the left. A complete set of the 16 speed-duration stimuli comprise a "replication" of the experiment. The order of presentation of the 16 stimuli was determined by means of a random number table for each replication. A sample of the stimulus tape is shown in Figure 1.

Subjects

The *Ss* were 10 male undergraduates at the Ohio State University. They were chosen from a list of students who had contacted the laboratory seeking employment and were paid one dollar per session. None had previous experience in tracking, motion prediction, or radar.

Procedure

Under the tracking condition, *S* traced the visible target in the slit with a pencil and continued tracing its predicted position until the trial ended. Under the monitoring condition, *S* simply watched the target with the pencil in his hand and then began tracing its predicted position after it disappeared. From the time that the target disappeared, the two conditions were identical.

The 10 *Ss* were randomly assigned to two groups. Group A began with the tracking condition and Group B with the monitoring condition. Each *S* participated in the experiment for 5 days, the first being devoted to instructions and two practice replications under the initial response condition. On the second and third days *S* ran 3 replications each day under his initial response condition. On the remaining 2 days he ran 3 replications per day under the opposite condition. No practice was given for the

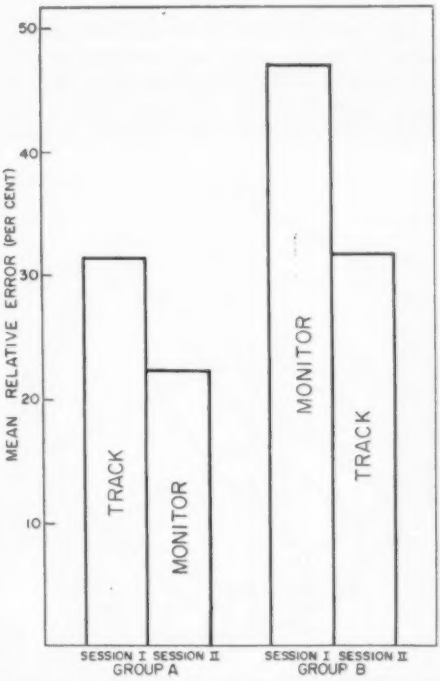

FIG. 2. Mean relative error by sessions and groups.

TABLE 1

LATIN SQUARE ANALYSIS

Source	*df*	*MS*	*F*
Conditions	1	79	2.32
Groups (order)	1	671	1.60
Sessions	1	616	18.1**
Subjects/Groups	8	420	12.4**
Error	8	34	
Total	19		

** $p < .01$.

transfer condition. An entire replication took about 8 minutes. After each replication *S* received a 5-minute rest period.

The responses were scored by extending the target line to the terminal point, and measuring to the closest millimeter the absolute distance from the terminal point to *S*'s terminal estimate. The mean of each *S*'s six scores under each speed-duration-condition combination was computed and formed the basis for the analysis. Thus each data point was the mean of six independent observations by each *S*.

RESULTS

The mean of each *S*'s six measurements in terms of absolute error was converted to relative error by dividing each mean by the distance traveled in the 9 seconds that *S* was making his predictions. For example, an absolute error of 9 mm. for a 4 mm/sec target was equal to 9 mm/36 mm or a 25% relative error.

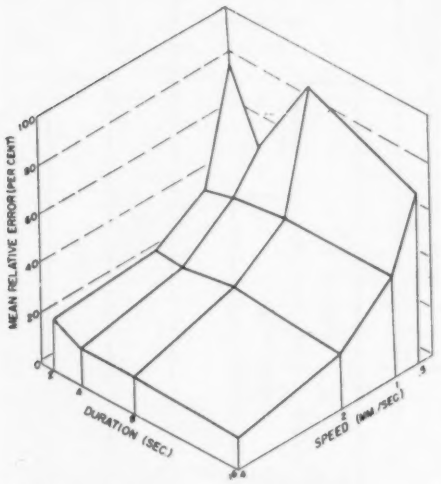

FIG. 3. Mean relative error as a function of duration and speed: Session I.

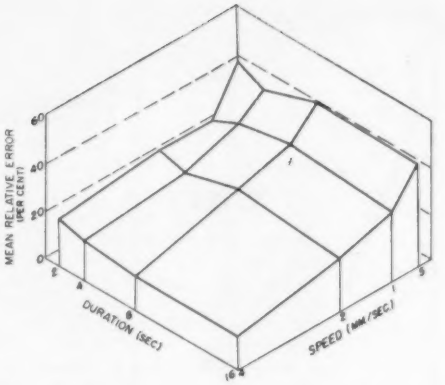

FIG. 4. Mean relative error as a function of duration and speed: Session II.

The mean relative error by groups and sessions (averaged across all 16 stimulus presentations) is shown in Figure 2. The data are analyzed by a 2×2 Latin square design as outlined by Grant (1948). The results of this analysis are shown in Table 1. These results indicate no difference between response conditions or order of presentation (groups), but only a significant learning effect from Session I to Session II and individual differences within groups.

TABLE 2

ANALYSIS OF VARIANCE FOR SESSION I
(First 6 trials)

Source	*df*	*MS*	*F*
Groups (G)	1	9,630	
Subjects/Groups	3	5,082	
Speeds (S)	3	26,911	14.6***
S × G	3	2,377	1.29
Error	24	1,846	
Durations (D)	3	807	1.77
D × G	3	1,547	3.19*
Error	24	485	
S × D	9	519	1.03
S × D × G	9	860	1.71
Error	72	502	
Total	159		

* $p < .05$.
*** $p < .001$.

TABLE 3

ANALYSIS OF VARIANCE FOR SESSION II
(Second 6 trials)

Source	*df*	*MS*	*F*
Groups	1	2,320	
Subjects/Groups	3	2,169	
Speeds	3	4,749	8.27***
S × G	3	1,526	2.66
Error	24	574	
Durations	3	101	
D × G	3	61	
Error	24	104	
S × D	9	82	
S × D × G	9	98	
Error	72	160	
Total	159		

*** $p < .001$.

Since the sessions dimension was the only main variable demonstrating statistical significance, the two sessions were analyzed separately for speed and duration effects. These analyses of variance took the form of a Lindquist Type VI design (Lindquist, 1953) with four speeds, four durations, and two response conditions. Each *S* performed under all of the 16 speed-duration combinations, but was nested in one of the response (tracking-monitoring) conditions. These results are displayed as three-dimensional plots in Figures 3 and 4. These plots indicate the marked effect of the speed variable and the lack of influence of duration of presentation. This is borne out in the analysis of variance presented in Tables 2 and 3. In both analyses, the speed variable was highly significant while the duration variable and the speed-duration interaction were not. Again response conditions failed to produce significant differences.

DISCUSSION

This study demonstrated the marked effect of target speed upon motion prediction accuracy with constant-rate targets. Previous investigations utilized static displays and failed to produce differences attributable to target speed. But on a static display speed of the initiating target is represented only by analogy, by varying the length of the trace line, or number of blips. It is highly questionable whether the subjects perceive this as speed at all, whereas with a dynamic display the subjects are exposed to a true sample of target motion.

An interesting result from this study is the lack of influence of duration of target presentation. It is surprising that the operator is

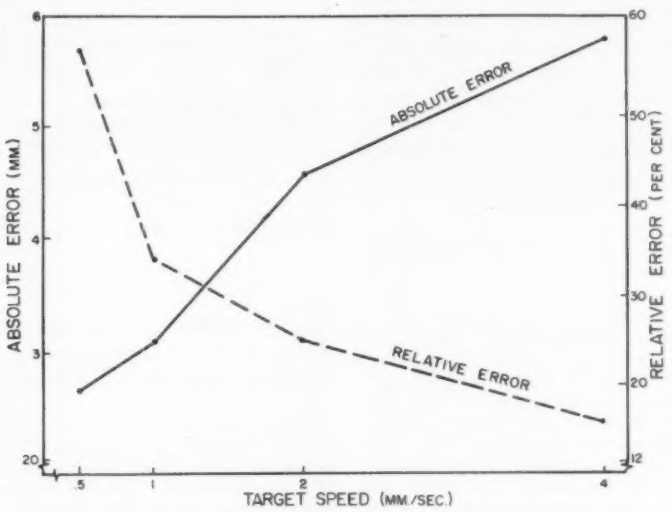

FIG. 5. Absolute error and relative error as a function of target speed.

able to make his prediction as accurately from a 2-second sample of target motion as from a 16-second sample. This result tends to support the view of Bowen and Woodhead (1953), who state: "As long as an observer has a minimum of information, he does not or cannot utilize any further information about the trace line" (p. 2). This has certain implications to system design where the operator or display mechanism must time-share. The results of this study would indicate that only a minimal amount of exposure of the target under control is necessary for accurate motion-prediction of constant rate targets.

The failure of the tracking-monitoring comparison to produce differences or transfer effects indicates that it makes little difference whether *S* receives his target information input actively or passively. Figure 2 indicates a trend toward a marked transfer effect from tracking to monitoring, which might have implications for training for a task where tracking is not possible. However, this effect failed to achieve statistical significance.

One brief remark on the performance measure employed in this study is advisable. The raw absolute error data were transformed to relative error, as it appeared to be more meaningful to express error in terms of the speed (and therefore distance traveled in the prediction interval). If the performance measure had been absolute error uncorrected for target speed, the opposite relationship of error to speed would have resulted as shown in Figure 5. This would imply that in application of these results to radar-like displays, the future position of high-speed targets would be most inaccurately predicted.

REFERENCES

BOWEN, H. M., & WOODHEAD, M. M. A prediction experiment. Cambridge: RAF Radar Unit, Psychology Laboratory, January 1953.

BOWEN, H. M., & WOODHEAD, M. M. Estimation of track targets after pre-view. *Canad. J. Psychol.*, 1955, **9**, 239–246.

GOTTSDANKER, R. M. The accuracy of prediction motion. *J. exp. Psychol.*, 1952, **43**, 26–36. (a)

GOTTSDANKER, R. M. Prediction motion with and without vision. *Amer. J. Psychol.* 1952, **63**, 533–543. (b)

GOTTSDANKER, R. M. A further study of prediction motion. *Amer. J. Psychol.*, 1955, **68**, 432–437.

GRANT, D. A. The Latin square principle in design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.

LINDQUIST, E. F. *The design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCGUIRE, J. C. Effects of traffic configurations on the accuracy of radar air traffic controller judgments. *USAF WADC tech. Rep.*, 1956, No. 56–73.

MANGLESDORF, J. E. Variables affecting the accuracy of collision judgments on radar-type displays. *USAF WADC tech. Rep.*, 1955, No. 55–462.

MANGLESDORF, J. E., & FITTS, P. M. Accuracy of joint extrapolation of two straight lines as a function of length of line. Paper read at Midwestern Psychological Association, Columbus, Ohio, 1954. (a)

MANGLESDORF, J. E. & FITTS, P. M. Accuracy of joint extrapolation of two aircraft trails as a function of length of trail. Unpublished report, Laboratory of Aviation Psychology, Columbus, Ohio, 1954. (b)

SCHIPPER, L. M., & VERSACE, J. Predictions of arrival sequences of simulated radar targets as a function of display size, target size, and target sharpness. *USAF WADC tech. Rep.*, 1956, No. 56–72.

STRATTON, G. M., MCCOMES, H. C., COOVER, J. E., & BAGBY, E. Psychological tests for selecting aviators. *J. exp. Psychol.*, 1920, **3**, 405–432.

(Received January 29, 1962)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 425–427

VIGILANCE PERFORMANCE AS A FUNCTION OF INTERPOLATED REST [1]

BRUCE O. BERGUM AND DONALD J. LEHR

United States Army Air Defense Human Research Unit, Fort Bliss, Texas

2 experiments were performed on the effects of interpolated rest upon monitoring performance at both high and low signal rates. Experiment I employed 2 groups of 20 Ss each; Experiment II employed 2 groups of 10 Ss each. One group of Ss worked on a light monitoring task for 3 30-min. periods separated by 10-min. rest periods. The 2nd group worked continuously for 90 min. on the same task. Experiment I employed 24 signals/hr; Experiment II employed 5 signals/hr. The results indicated a highly significant facilitation of detection performance as a result of interpolated rest at both signal rates and demonstrate the effectiveness of relatively brief rest intervals in maintaining high performance even with low signal rates.

The introduction of rest periods has consistently yielded improved vigilance detection performance over performance under nonrest conditions. Solandt and Partridge (1946), Saldanha (1956), and Jenkins (1958) have all reported various degrees of facilitation employing various lengths and frequencies of rest intervals. Probably the most dramatic effects were those reported by Mackworth (1950), who found that when he alternated half-hour watches with half-hour rest periods, uniform detection performance occurred, and in fact, performance after the rest period was even slightly improved over that for the first half-hour.

Of those studies reported, however, all have employed a relatively high signal presentation rate and none has been concerned with the extent to which the effects of interpolated rest can be generalized to tasks in which a relatively low signal rate is employed. The purpose of the present studies was to determine the extent to which such generalizations are valid.

EXPERIMENT I

This experiment tested the effects of interpolated rest on performance of a task employing a relatively high rate of signal presentation. The primary purpose of this study was to serve as a reference standard for Experiment II.

Method

Subjects. A total of 40 National Guard trainees from the Army Training Center, Fort Bliss, served as the subjects in this experiment.

Apparatus. Four isolation booths were employed in this study. Each booth was equipped with a circular panel, 13 inches in diameter, consisting of 20 ½-inch red lamps which illuminated in sequence at a rate of 12 rpm. A signal consisted of the failure of a lamp to illuminate in its normal sequence. The displays were mounted vertically at seated eye height on the rear walls of the booths and a small table was located directly beneath the display and adjacent to the wall. Room illumination was by a shaded 40-watt frosted lamp mounted above and behind the subject.

A preset program consisting of 24 signals per hour (Mackworth, 1950) caused a signal to be generated on the display and responses were made by depressing a hand-held pushbutton. Both signals and responses were automatically recorded on paper-tape recorders located in a central control area external to the booths. The control area and four booths were connected by a two-way intercommunication network.

Conditions. All subjects worked in isolation through three periods of approximately 30 minutes each, for a total of 90 minutes of monitoring. The 40 subjects were randomly assigned to two groups of 20 subjects each. The subjects in the first group worked continuously through the 90-minute period, while subjects in the second group were permitted a 10-minute rest period outside the test booths between the first and second, and second and third, monitoring periods. The response measure was percentage of correct detections.

[1] The research reported in this study was performed for the Human Resources Research Office of the George Washington University under contract to the Department of the Army. The opinions expressed in this report are solely those of the authors and do not necessarily reflect those of the sponsoring agency.

Results

The mean percentage of correct detections for both groups for each of the three time periods are presented in Table 1. The rest condition shows an almost constant high level of performance (approximately 93% correct detections overall), while the massed condition shows a marked and continuing performance decrement with progressive work periods.

A Kruskal-Wallis one-way analysis of variance for these data yielded an H of 5.44, $p < .01$. For this rate of signal presentation, rest periods yielded significantly superior detection performance.

EXPERIMENT II

This experiment tested the effect of rest intervals on detection performance in a task employing a relatively low signal presentation rate.

Method

Subjects. A total of 20 National Guard trainees from the Army Training Center, Fort Bliss, served as the subjects in this experiment.

Apparatus. The apparatus for this experiment was identical to that employed in Experiment I.

Conditions. The conditions for this experiment were identical to those employed in Experiment I with the exceptions that only 10 subjects served in each group and the signal rate was 6 signals per hour rather than the 24 signals per hour employed in Experiment I.

Results

The mean percentage of correct detections for both groups for each of the three test periods are presented in Table 2. The results are similar to those obtained in Experiment I. The rest condition shows an almost constant high level of performance over the full testing period, while the massed condition shows a marked performance decrement with progressive work periods.

TABLE 1

PERCENTAGE OF CORRECT DETECTIONS FOR REST AND CONTINUOUS GROUPS AT HIGH SIGNAL RATE

Condition	Time Period 1	2	3	Average
Rest	94%	94%	92%	93%
Continuous	92%	68%	73%	77%
Average	93%	81%	83%	

TABLE 2

PERCENTAGE OF CORRECT DETECTIONS FOR REST AND CONTINUOUS GROUPS AT LOW SIGNAL RATE

Condition	Time Period 1	2	3	Average
Rest	100%	100%	93%	98%
Continuous	97%	73%	70%	80%
Average	98%	87%	82%	

A Kruskal-Wallis one-way analysis of variance for these data yielded an H of 4.73, $p < .025$. As was the case in Experiment I, rest periods resulted in superior detection performance.

DISCUSSION

The introduction of rest intervals results in significantly superior detection performance over the entire monitoring period employed in these studies. In addition, the effect of this variable appears to be equally effective even when the signal presentation rate is reduced by a factor of four. This finding is of interest in view of the fact that several writers (Deese & Ormand, 1953; Jenkins, 1958; Kappauf & Powe, 1959; Nicely & Miller, 1957) have reported significantly poorer detection performance with lower signal rates. On this point, it is of special interest to note that the continuous group in Experiment I tended to perform at a slightly lower level than the continuous group in Experiment II, although the signal rate was higher for the former group. Comparison of these groups in terms of a Mann-Whitney U test, however, indicated no statistically significant difference between the groups. These results are very similar to those for the control groups in two experiments reported by Bergum and Lehr (1962). Signal rates of 6 and 24 per hour were employed in the latter study in conjunction with a paired-monitoring variable. The results showed almost identical control group performances at both rates. These results, along with those reported by Wherry and Webb (1951), who found no differences

in performance on an auditory task employing signal rates of 12 and 30 per hour, suggest that this variable is less general in character than has ordinarily been thought. It seems worth noting that those studies which have reported significant effects for signal rate have employed rates differing by a factor of from 4 to 60 between high and low rates, whereas in those studies in which no differences were demonstrated the rates differed by a factor no greater than four. It appears that the effects of this variable tend to break down at the lower signal rate ratios, suggesting a possible threshold effect for differential performance as a function of rate differences.

In terms of the rest variable, the present results indicate that interpolated rest is among the most powerful variables reported to date in this area. It is worth noting that Experiment I yielded results very similar to those found by Mackworth (1950) employing the same signal schedule, but with only one-third the amount of rest and suggests strongly that the minimum facilitative length of rest period may be well under the 10 minutes employed in the present studies.

REFERENCES

BERGUM, B., & LEHR, D. Vigilance performance as a function of paired monitoring. *J. appl. Psychol.*, 1962, **46**, 341–343.

DEESE, J., & ORMAND, E. Studies of detectability during continuous visual search. *USAF WADC tech. Rep.*, 1953, No. 53–8.

JENKINS, H. M. The effects of signal-rate on performance in visual monitoring. *Amer. J. Psychol.*, 1958, **71**, 647–661.

KAPPAUF, W. E., & POWE, W. E. Performance decrement on an audio-visual checking task. *J. exp. Psychol.*, 1959, **57**, 49–56.

MACKWORTH, N. H. Researches on the measurement of human performance. *Med. Res. Council spec. Rep.*, 1950, No. 268.

NICELY, P. E., & MILLER, G. A. Some effects of unequal spatial distribution of the detectability of radar targets. *J. exp. Psychol.*, 1957, **53**, 193–195.

SALDANHA, E. L. An investigation into the effects of prolonged and exacting visual work. *Med. Res. Council, Appl. Psychol. Res. Unit Rep.*, 1956, No. 243.

SOLANDT, D., & PARTRIDGE, D. M. Research on auditory problems presented by naval operations. *J. Canad. med. Serv.*, 1946, **3**, 323–329.

WHERRY, R. J., & WEBB, W. B. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1224–1236.

(Received February 2, 1962)

Journal of Applied Psychology
1962, Vol. 46, No. 6, 428–432

THE IDENTIFICATION OF TRAITS RELATED TO MANAGERIAL SUCCESS

JOHN A. HICKS

Aerojet-General Corporation, Azusa, California

AND JOICS B. STONE[1]

California State Polytechnic College

To evaluate the effectiveness of a test battery in discriminating between successful and unsuccessful managers, a comprehensive battery was administered to 76 supervisors of an autonomous, medium-sized operating division. Results from the Guilford-Zimmerman Aptitude Survey, California Test of Mental Maturity, and Structured-Objective (S-O) Rorschach Test were related to ratings of overall performance, promotability, and versatility by peers and superiors. The test battery showed significant correlations with peer ratings in all three of these areas (.46–.68), but showed a significant correlation (.44) only with the supervisors' evaluation of versatility. The S-O Rorschach Test proved to be the most effective test in the battery.

A great deal of effort has been invested in trying to identify the characteristics which make a successful manager. In attempting to do this, many test and personal history variables have been explored, often with very little success.

That much of the success of the manager is attributable to his personal characteristics and his ability to operate intuitively in a variety of situations has long seemed obvious to students of the area. Although many individuals have explained leadership and success in situational terms, there has been a continuing effort to identify characteristics or background factors which would account for this success. Rosen, Levinger, and Lippitt (1961) sketch six characteristics which seem to lend power to the individual in various situations. Kirkpatrick (1961) describes a series of personal history variables which have been used to effectively predict success in Chamber of Commerce executives.

If, as Porter (1961) has suggested in a recent study, the same traits are important in lower level managers as are important to the success of middle management personnel, then it should be possible to identify these traits across a sample which covers several levels of these managers. This would permit us to use larger samples of personnel to identify certain basic traits as a stepping stone to consideration of the difference between levels and specialties.

This study was a pilot project which included almost all of the supervisors and managers in an independent operation. The purpose of the study was to determine if a broad battery of tests covering aptitudes, temperament, and creativity could be used to identify certain basic characteristics for selection, promotion, and training purposes.

In a secondary sense, this study was designed to determine the economic feasibility of a testing program for this level of personnel and to delimit the areas for further exploration. A follow-up phase of this project which calls for a study of the relationships between personal history variables and the criteria of managerial success defined in this study is currently under way.

METHOD

Sample

Seventy-six managers, supervisors, shop foremen, and senior engineers (engineering supervisors) of a product division of a major manufacturer participated in this study. This included all of the supervisors within the division with the exception of a few who could not be scheduled even after repeated efforts to do so.

Test Battery

One of the major areas of exploration was the relationship between temperament and managerial success. To avoid the weaknesses of questionnaires

[1] Affiliated with Charles A. McKeand Company at time of this study.

and yet retain objective scoring and standardized norms, the Structured-Objective (S-O) Rorschach Test was used. This test is a multiple-choice modification of the Rorschach and provides measurement on 15 variables (Stone, 1958). The original validation included a large industrial population and norms have been established covering 46 occupational groupings (California Test Bureau, 1960).

The mental ability portion of the battery consisted of the California Test of Mental Maturity and four sections of the Guilford-Zimmerman Aptitude Survey: Verbal Comprehension (VC), General Reasoning (GR), Numerical, and Spatial Visualization (SV). Despite the lack of industrial data on the Aptitude Survey, it was used because it provided a factored aptitude measure for superior adults (Guilford & Zimmerman, 1956).

To measure the area of creativity the Executive test was used. The results from this test are not included in this report, however, since the scoring procedures and norms are still in the experimental stage.

Criteria

It was concluded early in the study that any measure of managerial performance would be heavily weighted by the opinions of those individuals with and for whom each manager worked. It was decided, therefore, to obtain peer ratings and supervisory ratings on three areas: overall performance, promotability, and versatility. The test battery was then related to each of the six resulting scales.

To obtain the peer ratings the participants were grouped into 10 groups consisting of 3–13 individuals. The individuals in each of these groups had substantial personal knowledge of the work of the other individuals, were at a comparable level, and had comparable professional backgrounds.

In order to provide some stability to the supervisors' ratings, two individuals in a superior capacity were asked to rate each participant. For this reason, not all of the raters had direct supervisory experience with the subordinate they rated, but all of the raters were in a position to review much information about the ratee's work.

The rating form was designed so that all individuals were rated on each trait before the rater proceeded to the next trait. The raters were instructed to make as much discrimination between the individuals as they could, keeping in mind the full range of alternatives available. When the ratings were received and compiled, each of the peer rating groups was compared to the other groups in terms of the mean and range of ratings given. These were similar enough so that no correction was considered necessary. The reliability of the peer ratings was .56 while that for supervisors was .52.

Procedure

After the purpose and procedure of the program were explained to the participating supervisors, they were given the rating forms for their peers and subordinates to be completed at the rater's convenience. As soon as all of the complete ratings had been received, the individuals were tested in three groups.

When the results had been tabulated and profiles made, each individual received a summary of his results. The testing results and their implications were discussed broadly in a group session. This was followed by private interviews for those who wished to discuss their results in detail.

To minimize the threat which these managers would feel from the testing program they were assured that (*a*) the results of the peer and supervisory ratings would be used only for purposes of validating the test battery and that complete data would not be made available to management; and (*b*) no data, either from the tests or the ratings, which would be detrimental to the individuals' careers would be released to management. In this regard they were told that division management would know who the outstanding individuals were on the ratings and on the test battery and that division management would receive statistical data showing the overall relationships.

Analytic Method

To facilitate the analysis of the data, each of the six sets of ratings were broken into the top 27%, middle 46%, and bottom 27%. A preliminary analysis was made to determine which test variables showed the greatest discrimination between the top and bottom groups on the six sets of ratings. Correlation coefficients were then calculated for these variables and a multiple correlation computed using them. Finally, by using the Taylor-Russell tables and estimates of turnover cost it was possible to prepare charts which showed the break-even points for the testing program under various conditions and the amount of saving which might be expected.

RESULTS

Table 1 shows the intercorrelations between the major dimensions of the criteria. Two basic conclusions may be drawn from this table: (*a*) there is a substantial amount of common variance between the three areas of job success, the smallest correlation being .70 between promotability and versatility;

TABLE 1

INTERCORRELATIONS BETWEEN CRITERIA

Criterion	Promotability	Versatility	Between peers and supervisors
Overall performance	.87	.74	.51
Promotability		.70	.59
Versatility			.59

TABLE 2

CORRELATIONS BETWEEN SELECTED TEST VARIABLES AND CRITERIA

Criterion	S-O Rorschach Test										Aptitude Survey			
	W	*D*	*Dd*	*F*	*M*	*FM*	*FC*	*A*	*H*	*P*	VC	GR	SV	*R*
Overall performance														
Peer	.15		−.52**	−.20	.48**	−.04		−.04	.39**		.15	.10	.00	.68**
Supervisor	.08		−.10	−.16	.13	−.04		−.12	.18		.18	.10	.23*	.30
Promotability														
Peer	.25*	−.12	−.12		.06	−.35**		−.15	.29**	.04	.15		.06	.55**
Supervisor	.25*	−.10	−.08		.25*	.11		−.04	.13	.20	.17		.11	.34
Versatility														
Peer	.06		−.38**		.14	−.21	.04		.25*					.46**
Supervisor	.04		−.29**		.17	.00	−.12		.08					.44**

* $p \leq .05$.
** $p \leq .01$.

(*b*) although the ratings of peers and supervisors are significantly correlated the relationship is not a large one, ranging from .51 to .59. The peers and supervisors were obviously using somewhat different standards in evaluating these individuals.

The correlations between the criteria and selected test variables are listed in Table 2. The last column contains the multiple correlation coefficients for each of the criteria.

The four S-O Rorschach Test subscores which showed the greatest relationship across the criteria are *W*, *Dd*, *M*, and *H*. These are defined by the S-O Rorschach Test manual as: Theoretical (*W*), Pedantic (*Dd*), Activity Potential (*M*), and Human Relationships (*H*). The Aptitude Survey variables showed only a small relationship to the criteria.

Certain of the test variables showed a pattern of relationship among the criteria. *H* was significantly correlated with all three of the peer scales. *W* was significant for both of the promotability scales and *Dd* for the versatility scales.

The test battery proved much more effective in predicting the peer ratings than it did in predicting the supervisors' ratings, as may be seen from a comparison of the multiple *R*s. On overall performance the test battery correlated .68 for the peer's ratings as contrasted with .30 for the supervisors' ratings, a difference significant at the 1% level. On promotability the difference (.55–.34) was significant at the 5% level, and on versatility there was no significant difference. Since there is no significant difference in the reliability of the two sets of ratings, these data suggest that each group sees significant dimensions of performance which the other does not.

To estimate shrinkage which would occur in applying these results to a new sample, Wherry's shrinkage formula was used. The corrected coefficients range from .35 to .63 for the three peer rating scales, all significant above the 5% level. For the supervisor's rating of versatility, the corrected correlation was .30, also significant at the 5% level.

Another facet of this problem was the relationship between the reliability of the criteria and the correlation coefficients. Using the formula for the correction for attenuation, the expected correlation was estimated using the assumption that the obtained coefficient was the true correlation. Under this assumption the same four coefficients ranged from .27 to .43. All of these except the coefficient for the peer rating of versatility (.29) were significant at the 5% level.

The conclusions about the validity of the test battery were not changed, then, using either an estimate of shrinkage or correction for attenuation to be expected from the criteria and test reliabilities. To answer the question of shrinkage conclusively, however, a cross-validation would be necessary.

DISCUSSION

It appears, from this study, that managerial success can be predicted to a significant degree using current testing instruments.

Managers in this study represent a wide range of skills and background and have in common only the fact that they are operating as managers in a technical organization.

This leads to the second conclusion that, in this organization at least, there are certain basic characteristics which the successful managers possess regardless of their areas of specialization. We would expect that there are measureable differences between managers in widely divergent fields such as engineering, production, and administration; however, until one has established the things which they have in common these other questions appear secondary.

Another interesting result of this study was that peer evaluations of overall performance and promotability appeared to be more effective in prediction than evaluations made by the superiors, with differences significant at the 1% and 5% levels, respectively. This raises questions about the total evaluation process in an organization. One reasonable hypothesis might be that the man's associates know him much better than his superiors do, since a man is always trying to impress his superiors. Some of the incidental data which were collected in this study indicated that this was the case for many of these individuals.

On the other hand, it is equally reasonable to assume that the supervisor has a more realistic view of the man in terms of the managerial requirements of the organization. The fact that none of the test variables predicted well the supervisors' ratings, however, suggests that the results may have simply been a matter of inaccuracy in the ratings of the supervisors.

This result posed another problem. In applying test results it was necessary to know which types of job success to predict, i.e., success in the eyes of his peers or his supervisor. The decision was made that in the early stages of the application of this data the supervisor hiring the man would simply be given a statement of the probabilities that this man would be highly regarded by him and by the man's peers, permitting the supervisor to assign the relative weights.

Mental aptitudes did not show a large relationship to managerial success in this study, a fact which has been pointed out in many earlier studies. This has been interpreted as the result of sample attenuation through preselection on education and career achievement.

It is interesting to compare this result with Porter's (1961) study in which both bottom and middle managers ranked the description, "intelligent," as one of the most important traits for managerial success. Either managers are describing other traits as intelligence or they simply believe that people like themselves must be intelligent.

The success of the S-O Rorschach Test in identifying certain temperament characteristics which related to managerial success suggests that it may be an instrument which would be particularly applicable to industrial settings. As a multiple-choice test which cannot be faked, it avoids many of the problems which questionnaires have.

To the extent that the constructs are valid and they follow basic Rorschach theory, we can describe the successful manager in this organization as an individual who shows a great deal of emotional strength; who does not structure or organize his activities to the extent that the less successful managers do; and who tends to view things from a broad, theoretical point of view, avoiding overinvolvement in detail. Much work needs to be done to define the responsibilities of the manager before we can be sure what these interpretations mean in such an industrial setting.

The finding on emotional strength seems to bear out an earlier conclusion in a study of union business agents (Rosen & Rosen, 1957). Using the MMPI, they found that one of the scales which seemed to discriminate between successful and unsuccessful agents was the Ego Strength scale. It would seem in terms of face validity that much of the same factor is being measured in these two tests. Emotional resilience and the ability to generate and sustain directed activity are indicated in both scales.

Another problem for future research is the discrepancy between peer and supervisory ratings. Should this prove to be a common phenomenon, a study of its causes and the relative validity of the two viewpoints would do much toward clarifying what true job

effectiveness is and what types of cues are used to measure it.

REFERENCES

CALIFORNIA TEST BUREAU. *Research report: Industrial norming project.* Los Angeles: CTB, 1960.

GUILFORD, J. P., & ZIMMERMAN, W. S. *Guilford-Zimmerman Aptitude Survey manual.* Los Angeles: Sheridan Supply, 1956.

KIRKPATRICK, J. J. How to select a Chamber of Commerce executive. *J. Amer. Chamber Comm. Exec.*, 1961, **10**.

PORTER, L. W. Perceived trait requirements in bottom and middle management jobs. *J. appl. Psychol.*, 1961, **45**, 232–236.

ROSEN, S., LEVINGER, G., & LIPPITT, R. Perceived sources of social power. *J. abnorm. soc. Psychol.*, 1961, **62**, 439–441.

ROSEN, H., & ROSEN, R. A. H. Personality variables and role in a union business agent group. *J. appl. Psychol.*, 1957, **41**, 131–136.

STONE, J. B. *S-O Rorschach Test manual.* Los Angeles: California Test Bureau, 1958.

(Received February 12, 1962)

MANAGEMENT SCIENCE

QUARTERLY JOURNAL OF THE INSTITUTE OF MANAGEMENT SCIENCES

Make TIMS your information center for new ideas in the Management Sciences

For membership and subscription information write to

THE INSTITUTE OF MANAGEMENT SCIENCES
Box 626
Ann Arbor, Michigan

Journal of Engineering Psychology

Edited by LEE W. COZAN

Original experimental research in human factors engineering: Systems, information displays, controls and layout of workplaces, skilled performance, monitoring and vigilance, and unusual environment and stress.

Published quarterly beginning January 1962

Annual Subscription: $8.00 (Overseas $9.00) **$2.50 per issue**

Send Orders and Manuscripts to

ELIAS PUBLICATIONS

P. O. Box 662, Washington 4, D. C.